**Высоцкая Е. В.**

# МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ В ПРИКЛАДНЫХ МЕДИЦИНСКИХ ИССЛЕДОВАНИЯХ

**Харьков - 2013**

**Высоцкая Е.В.** Многомерный статистический анализ в прикладных медицинских исследованиях: монография. – Харьков, 2013. – 136 с.

В монографии на реальных медицинских примерах рассмотрены сложные методы анализа данных с использованием пакетов прикладных программ STATISTICA и SPSS. Особое внимание уделено интерпретации результатов.

Книга адресована студентам, аспирантам и специалистам, применяющим методы прикладной статистики в медицине.

Ил. 72. Табл. 17. Библиогр. наимен. 63.

**Рецензенты:**

С.М. Злепко, д.т.н., профессор, заведующий кафедрой ПМБА Винницкого национального технического университета.

Л.И. Рак, д.мед.н., ведущий научный сотрудник отделения педиатрии и реабилитации ГУ «Институт охраны здоровья детей и подростков АМН Украины».

И.Ю. Шубин, к.т.н., доцент, профессор кафедры ПИ Харьковского национального университета радиоэлектроники.

**ISBN 978-966-97289-1-3**

**СОДЕРЖАНИЕ**

# Введение

В начале 90-х годов было введено в обиход специалистами понятие «доказательная медицина», концепция которой подразумевает добросовестное, точное и осмысленное использование лучших результатов клинических исследований для выбора лечения конкретного больного. Подобный подход позволяет снизить уровень врачебных ошибок, облегчить процесс принятия решения для практических врачей и администрации лечебных учреждений, а также уменьшить расходы на здравоохранение и использовать сэкономленные средства для претворения в жизнь социально ориентированных медицинских проектов [1].

Согласуясь с принципами доказательной медицины, врач обязан критически подходить к решению возникающих перед ним проблем и, как правило, самостоятельно предпринимать те или иные действия. Современную медицину невозможно представить себе без использования точных и надежных методов анализа.

Обоснованность принимаемых решений в настоящее время в мировой практике оценивается с помощью статистических методов.

Однако проблема состоит в том, что в большинстве случаев объекты медицинских исследований являются сложными медицинскими системами с множеством элементов в своем составе и связей между ними, определяющих внутреннюю структуру системы. Такие системы (объекты, процессы) формируются и функционируют под воздействием многочисленных изменчивых и взаимодействующих между собой факторов среды [2 – 4]. Следовательно, для применения многомерного статистического анализа имеются все основания. На сегодняшний день разработана теория и практика адекватного кодирования медицинских данных. Современные технические и программные средства информатизации обеспечивают возможность создания баз данных в объемах, достаточных для получения достоверных результатов. Наличие пакетов прикладных программ по математико-статистической обработке и анализу данных позволяют применять многомерные математические методы и процедуры специалистами-медиками. Создаваемые на этой основе многомерные математико-статистические модели сложных медицинских систем позволяют перейти от интуитивных представлений

врача о медицинских процессах и системах к математически обоснованному количественному анализу.

Современные компьютерные программы призваны оказать эффективную помощь исследователю, однако, при работе с компьютером возникают проблемы интерпретации полученных результатов.

Наука означает знание. Понимание и объяснение являются непрерывными процессами, получающими развитие по мере того, как выявляются все новые и новые взаимосвязи между различными научными законами.

В монографии просто и наглядно изложены результаты научных исследований, проведенных автором совместно с врачами-исследователями.

На реальных медицинских примерах статистическими методами (множественный регрессионный анализ, дискриминантный анализ, анализ временных рядов) была исследована структура медицинских данных. Исследование и анализ данных позволили решить задачу идентификации состояния больных и прогнозирования. В качестве инструментария исследований использованы пакеты прикладных программ STATISTICA 10, SPSS 21. Особое внимание уделено интерпретации полученных результатов.

В первой и второй главах представлены результаты применения дискриминантного анализа при определении хронической сердечной недостаточности (ХСН) и дисфункции миокарда у подростков. Для каждого из заболеваний выделено небольшое количество клинико-лабораторных показателей, по которым можно с высокой достоверностью определить степень тяжести состояния больного. Показана возможность и целесообразность использования дискриминантного анализа для идентификации состояния больных с указанными заболеваниями.

В третьей главе рассмотрено применение бинарной логистической регрессии для определения сердечно-сосудистого риска у пациентов с почечной недостаточностью. Были изучены факторы риска сердечно-сосудистых заболеваний, среди которых выделены наиболее значимые, оказывающих наиболее значительное влияние на прогрессирование хронической ишемической болезни сердца в виде ишемической дилятационной кардиомиопатии.

Четвертая глава посвящена вопросам прогнозирования заболеваемости расстройствами психики и поведения непсихотического характера среди сельского населения Украины. Посредством анализа временного ряда построены модели ряда, по которым можно делать прогноз интересующего показателя на будущий временной интервал в различных регионах Украины.

Книга рассчитана на широкий круг читателей – студентов, аспирантов и специалистов, использующих анализ данных в медицинских исследованиях.

Результаты анализа данных и построенные вероятностно-статистические модели могут быть использованы при принятии медицинских решений по стратегии и тактике лечения больных.

# ГЛАВА 1. ОПРЕДЕЛЕНИЕ РАННЕЙ СТАДИИ ХРОНИЧЕСКОЙ СЕРДЕЧНОЙ НЕДОСТАТОЧНОСТИ У ПОДРОСТКОВ С ИСПОЛЬЗОВАНИЕМ SPSS

Изучение связей основных клинических параметров и их комбинаций является важнейшей задачей теоретической и практической медицины. В настоящее время медицина все чаще обращается в своих исследованиях к достижениям математических наук. Не вызывает сомнения тот факт, что статистическое моделирование основных жизненных функций в норме и патологии имеет большое теоретическое и практическое значение. Оно позволяет объяснить суть патологического процесса, правильно оценить прогноз заболевания и разработать эффективные лечебно - профилактические мероприятия.

Рост сердечно-сосудистых заболеваний в мировом масштабе и их «омоложение» в последнее время является общеизвестным фактом. Большинство проблем общества на современном этапе развития тем или иным образом связано с демографическим кризисом и инвалидизацией лиц трудоспособного возраста, ростом числа больных с различными нарушениями в работе сердечно-сосудистой системы и снижением продолжительности жизни.

В настоящее время профилактика формирования и прогрессирования ХСН является приоритетным направлением кардиологии. В педиатрической практике большой удельный вес занимают заболевания сердца, которые включают врожденные аномалии (врожденные пороки сердца, множественные малые структурные аномалии сердца) и приобретенные патологии (кардиты, вторичные кардиомиопатии и т.д.), на фоне которых возможно формирование ХСН уже в раннем возрасте [5 – 6].

Как известно, развитие ХСН является патофизиологически необратимым процессом, поскольку механизмы ее формирования, выполняя адаптационную функцию на начальных этапах, постепенно становятся дезадаптационными факторами, приводящими к прогрессированию процесса [7]. Основные звенья данного патологического процесса включают активацию нейрогуморальных систем регуляции, системы цитокинов и оксидативного стресса, ремоделирование миокарда, которые тесно связаны между собой. В то же время, на ранних

этапах при отсутствии выраженных клинических проявлений (одышка, отечный синдром) диагностика ХСН затруднена.

Наряду с общеизвестными медицинскими методами диагностики ХСН и определения тенденции развития систолической дисфункции миокарда все чаще исследователи обращаются к математическому описанию данных процессов.

В детерминированных моделях входные значения моделируемого процесса представляются в виде аналитических закономерностей, для которых может быть получен единственный результат. Эти модели могут отображать как детерминированную, так и вероятностную системы, но при этом описание последней бывает неполным, так как модель в таких случаях является упрощенной [8].

Стохастические модели, в которых информативные признаки, характеризующие состояние моделируемого объекта, представлены случайными величинами и связаны стохастическими (случайными, нерегулярными) зависимостями, позволяют решить эту проблему.

Так как параметры, характеризующие систолическую дисфункцию миокарда, могут принимать одно из множества значений, причем появление значения до измерения нельзя точно предсказать, то для разработки диагностической модели систолической дисфункции миокарда у подростков с учетом изменений морфофункциональных характеристик сердца, показателей систем нейрогуморальной регуляции, иммуновоспалительной активации и свободнорадикальных процессов необходимо применять стохастические методы моделирования.

Дискриминантный анализ дает возможность быстро и качественно классифицировать исследуемые объекты и синтезировать адекватную линейную математическую модель. Данный метод математического анализа имеет такие преимущества: учитывает вариабельность используемых в моделировании параметров, рассматривает совокупность всех клинических и параклинических показателей, взятых со своими коэффициентами, которые отражают удельный вес влияния каждого показателя на постановку диагноза [9 – 11].

Метод диагностики хронической сердечной недостаточности у подростков на ранней стадии развития включает в себя ряд этапов.

**Этап 1.** Сбор информации о состоянии организма пациента, включающий определение морфофункциональных характеристик сердца, путем проведения эхокардиографии в покое и после физической нагрузки, показателей систем нейрогуморальной регуляции, аутоиммунным активации и свободнорадикальных процессов.

Были проанализированы данные 159 подростков с заболеваниями сердечно-сосудистой системы. Все пациенты по клиническим критериям были разделены на 3 группы: 1-я группа - 27 подростков с заболеваниями сердечно-сосудистой системы без признаков ХСН (17%); 2-я группа - 65 подростков с заболеваниями сердца и начальной стадией ХСН (40,9%); 3-я группа - контрольная, которую составили 67 практически здоровых сверстников (42,1%).

Так как наименьший процент составили подростки с заболеваниями сердечно-сосудистой системы без сердечной недостаточности, то объем выборки для создания математической модели определялся с учетом этого, наименьшего показателя.

Все пациенты рассматривались как совокупность объектов с варьирующими количественными и качественными характеристиками, на основании которых определялась группа, к которой принадлежит объект. Это позволило предсказать группы (наличие / отсутствие сердечной недостаточности) для новых объектов из той же совокупности.

Метод дискриминантного анализа, как и вся группа методов, основанных на наблюдениях, предполагает использование определенной терминологии. Классифицируются объекты, а различные группы, к которым необходимо эти объекты относить, чаще всего называются классами. Алгоритм отнесения объектов к разным классам называется решающим правилом.

Нами рассматривались две выборки, по одной из которых строили решающее правило (обучающая), а по второй проверяли качество работы этого правила (контрольная).

**Этап 2.** Кодирование качественных и количественных показателей.

Для определения значимых для формирования диагноза свойств, было проанализировано 73 показателя. Все показатели были закодированы и поставлены в соответствие 73-мерному вектору, который учитывает

отсутствие, наличие, направленность и величину каждого показателя. Кодировочная таблица представлена в табл. А.1 (Приложение А).

Ни одна переменная не была линейной комбинацией других переменных. Т.е. мы не пользовались суммой переменных или их средним арифметическим совместно с самими переменными.

Математическая обработка результатов проводилась с использованием и пакета прикладных программ SPSS 21 для Windows .

Для ввода всех закодированных данных в компьютер, после запуска SPSS для Windows, с помощью двойного нажатия левой кнопки мыши на значке SPSS открывается редактор данных SPSS (рис. 1.1).

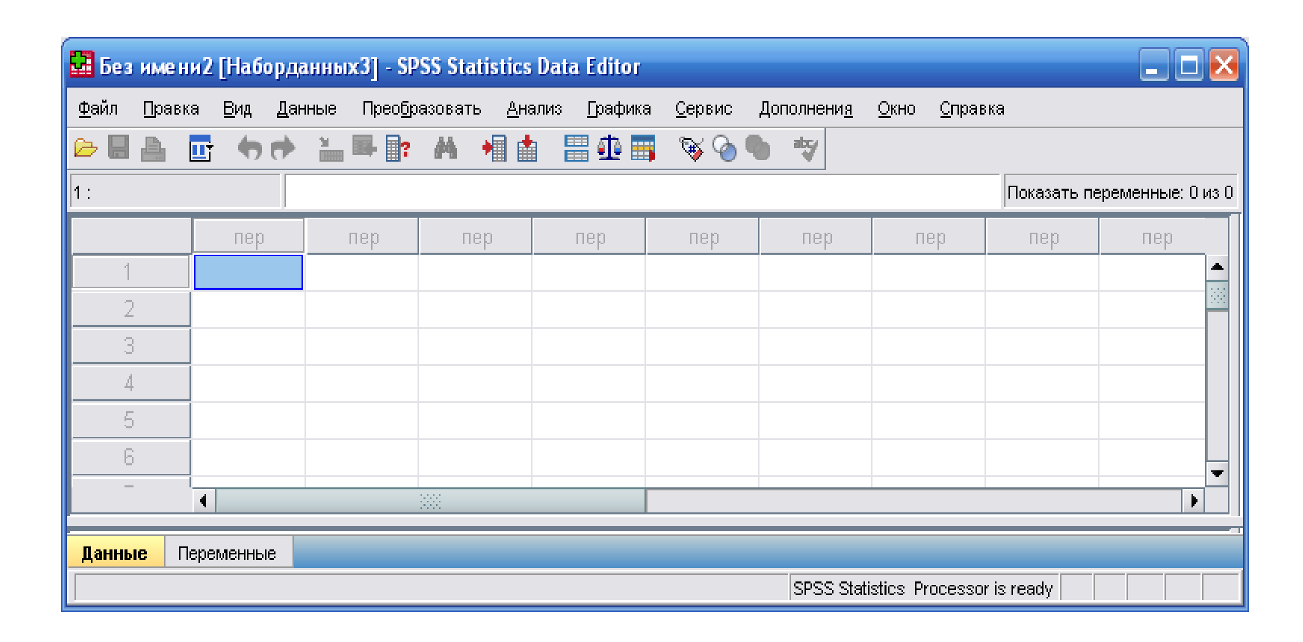



Рисунок 1.1 – Редактор данных: просмотр данных

В редакторе данных двойным нажатием на ячейках с надписью «пер» перешли в режим просмотра переменных, который обеспечивает редактор данных (рис. 1.2). Здесь последовательно, строка за строкой определялись необходимые переменные. Для этого двойным нажатием в таких полях, например, как «Имя», «Тип», «Ширина» и т.д., задавалась вся необходимая информация для вводимых переменных.

При кодировании качественной переменной sn, по которой мы группировали исследуемых пациентов, учли что этот показатель может принимать три значения: 1- «Без сердечной недостаточности», 2- «1-я стадия сердечной недостаточности», 3 – «Контрольная группа».

После открытия диалогового окна «Метки значений» (рис. 1.3) метки значений были определили следующим образом:

1) вначале ввели в поле «Значение» число «1» и нажали клавишу <Tab>;

ХСН (eeeee).sav [Наборданных1] - SPSS Statistics Data Editor

Файл Правка Вид Данные Преобразовать Анализ Графика Сервис Дополнения Окно Справка

| | Имя | Тип | Ширина | Десятич... | Метка | Значения | Пропуски | Столбцы | Выравнивание | Шкала |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | diagnoz | Числовая | 2 | 0 | Диагноз | {1, Миокар... | Нет | 8 | По право... | Номинальная |
| 2 | Diagnoz1 | Числовая | 8 | 0 | Диагноз | {1, Миокар... | Нет | 8 | По право... | Количестве... |
| 3 | sn | Числовая | 8 | 0 | Сердечная не... | {1, Без сер... | Нет | 8 | По право... | Количестве... |
| 4 | pol | Числовая | 2 | 0 | ПОЛ | {1, м}... | Нет | 8 | По право... | Номинальная |
| 5 | vozrast | Числовая | 3 | 0 | ВОЗРАСТ, года | Нет | Нет | 8 | По право... | Номинальная |
| 6 | ekg_chss | Числовая | 3 | 0 | ЧСС, ударов в... | Нет | Нет | 8 | По право... | Количестве... |

Данные Переменные

SPSS Statistics Processor is ready

Рисунок 1.2 – Диалоговое окно редактирования переменных

2) ввели в поле «Метка» текст «Без сердечной недостаточности»;

3) нажали на кнопке «Добавить», после чего метка значения была добавлена в список;

4) повторили эти действия для значений «2» — «1-я стадия сердечной недостаточности» и «3» — «Контрольная группа». Максимально допустимая длина метки значения составляет 60 знаков.

Таким образом, закодировали все качественные показатели.

**Этап 3.** Расчет необходимого объема выборки пациентов.

С учетом предельной ошибки выборки $\Delta_{\omega} = 5 - 6\%$. При надежности γ = 0,95 и рассчитанной предельной погрешности необходимый объем выборки составляет:

$$n_{\omega} = \frac{t^2 \cdot \omega \cdot (1-\omega)}{\Delta_{\omega}^2} = 148 \div 214 \text{ человек,}$$

где t – значение коэффициента доверия Стьюдента при заданном значении γ; ω – больные с заболеваниями сердца без сердечной недостаточности, %.

**Этап 4.** Описательный анализ данных.

На этом этапе определяли имеет ли каждая переменная нормальное распределение при фиксированных остальных. Если экспериментальные данные не вполне удовлетворяют этим предположениям, то статистические выводы не будут точным отражением реальности.

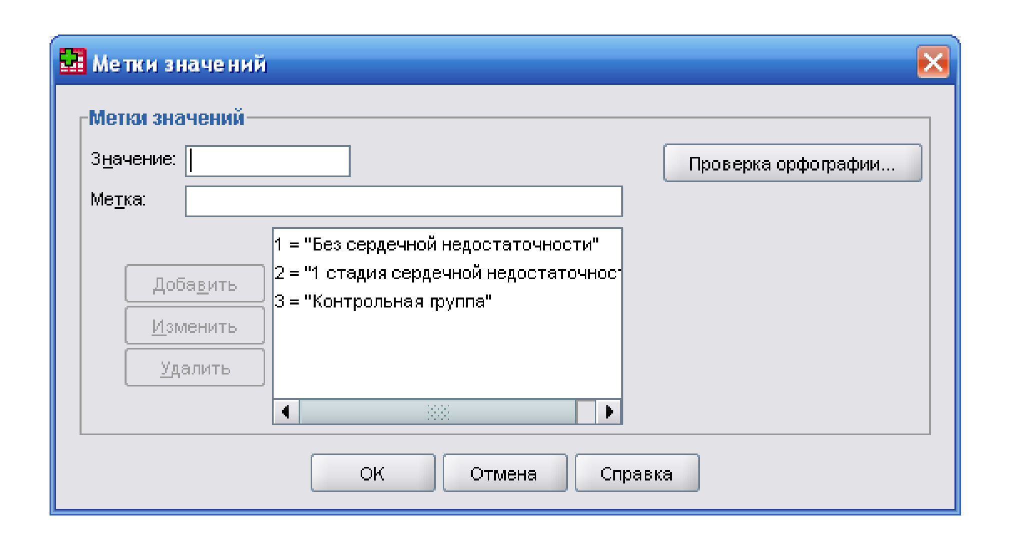



Рисунок 1.3 – Диалоговое окно «Метки значений»

Выбрали в меню «Анализ» пункт «Классификация» → «Дискриминантный анализ» (рис. 1.4).

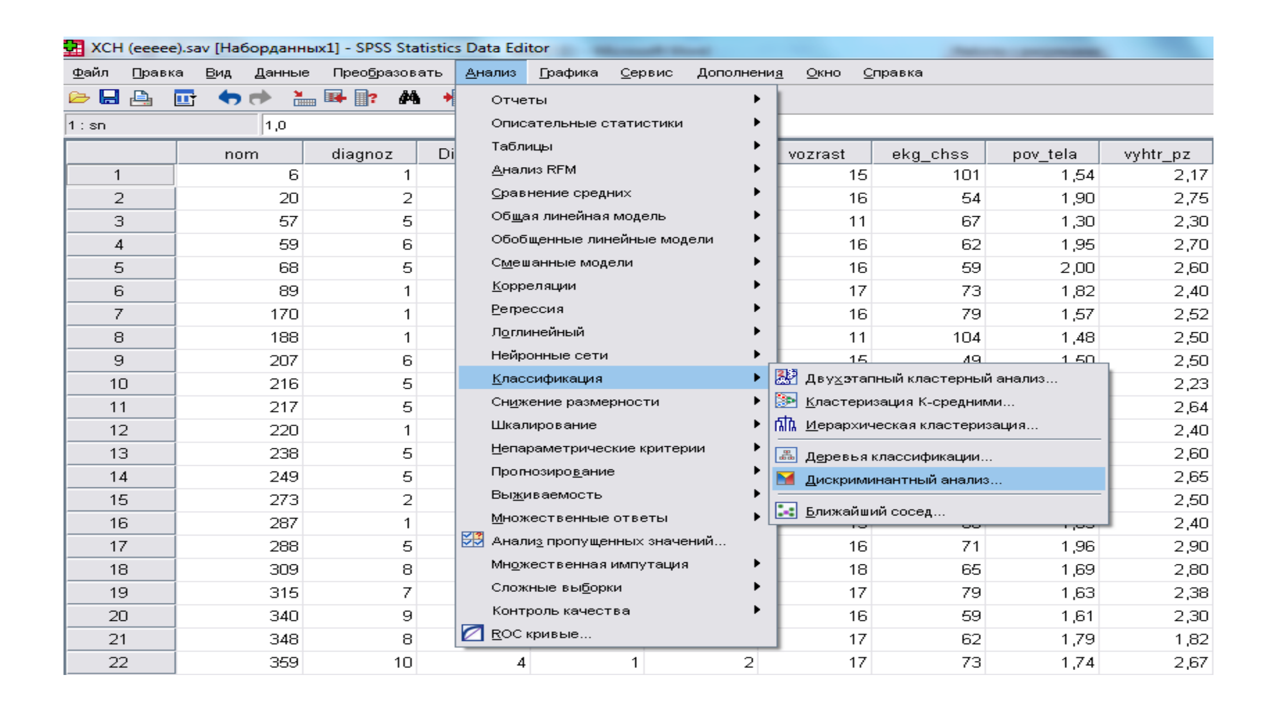



Рисунок 1.4 – Выбор метода анализа данных

После открытия диалогового окна «Дискриминантный анализ» (рис. 1.5) единственную зависимую переменную sn поместили в поле, предназначенное для группирующих переменных (рис. 1.6).

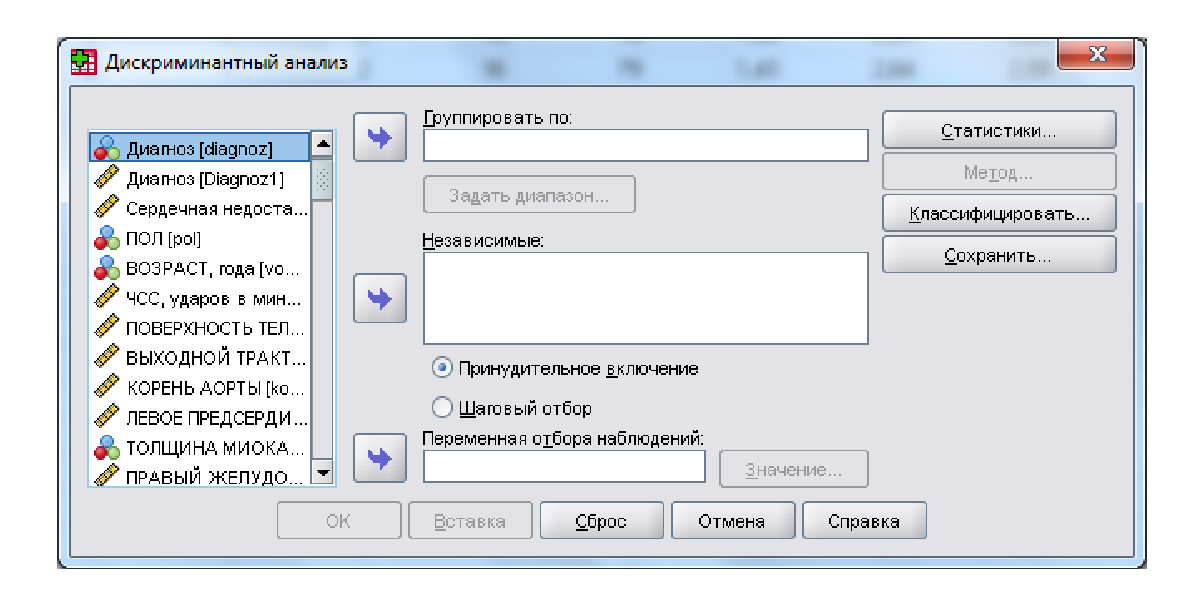



Рисунок 1.5 – Диалоговое окно «Дискриминантный анализ»

Нажатием на кнопке «Задать диапазон» ввели минимальное и максимальное значение этой переменной: 1 и 3 соответственно, так как данная переменная имеет три уровня (рис. 1.7). Затем нажимали на кнопку «Продолжить».

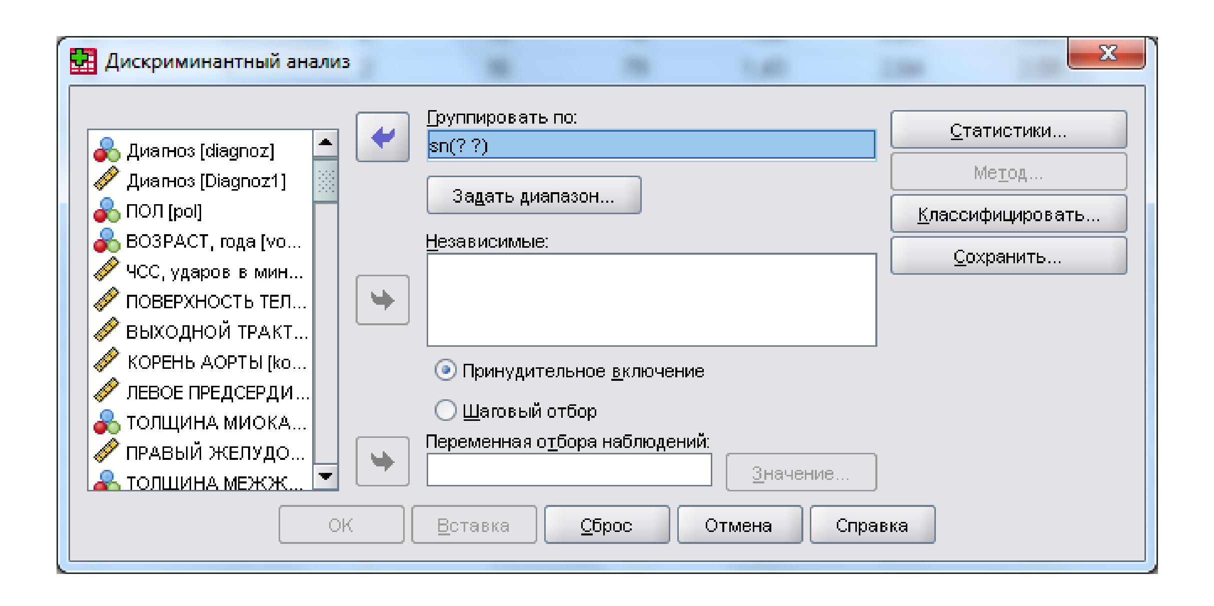



Рисунок 1.6 – Диалоговое окно «Дискриминантный анализ».
Выбор переменных

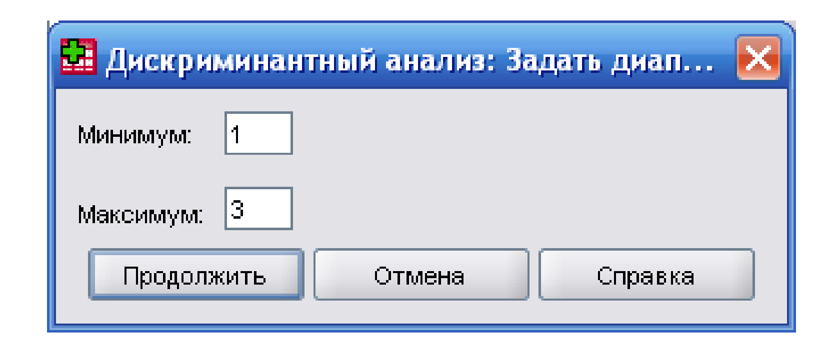



Рисунок 1.7 – Диалоговое окно «Дискриминантный анализ: Задать диапазон»

Всем остальным переменным присвоили статус независимых переменных, для этого перенесли их в поле «Независимые». Далее выбрали метод «Шаговый отбор» (рис. 1.8), который рекомендуется при наличии большого количества независимых переменных.

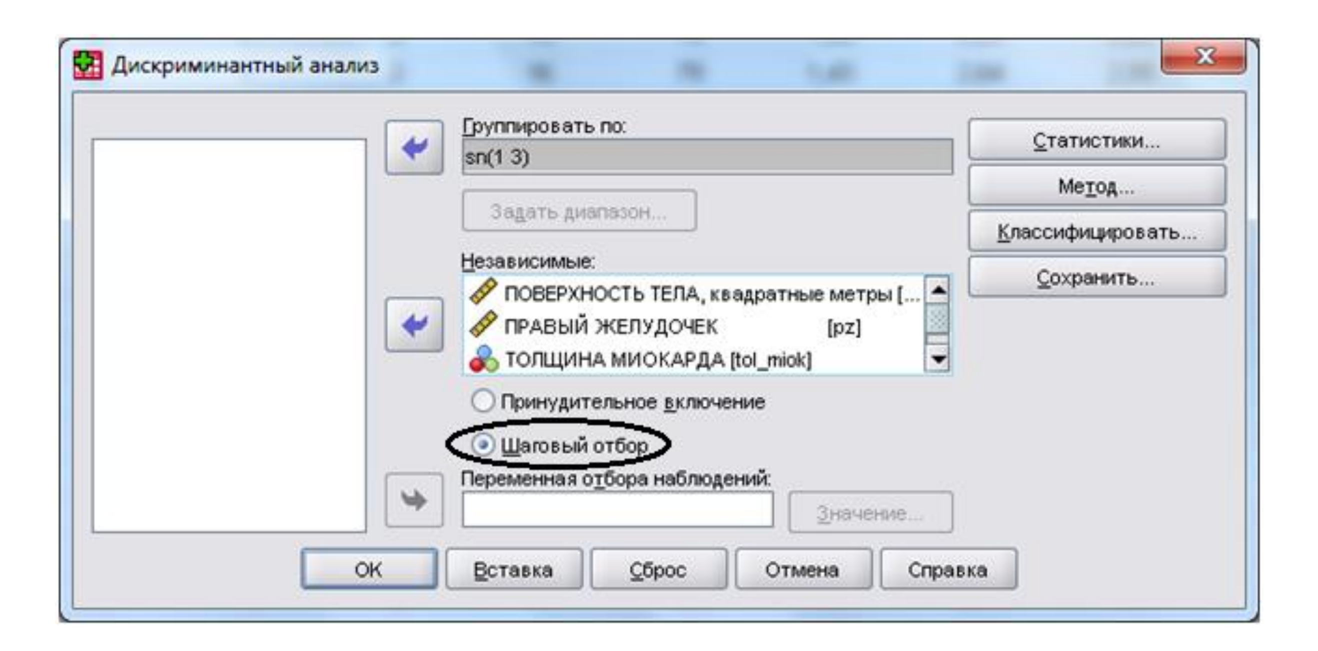



Рисунок 1.8 – Диалоговое окно «Дискриминантный анализ»

После нажатия на кнопку «Статистики…» активировали опции: «Средние», «Однофакторный дисперсионный анализ», «Фишера» и другие (рис. 1.9).

При выборе «Среднее» мы получили средние значения и стандартные отклонения для каждой переменной каждой группы и всей выборки в целом.

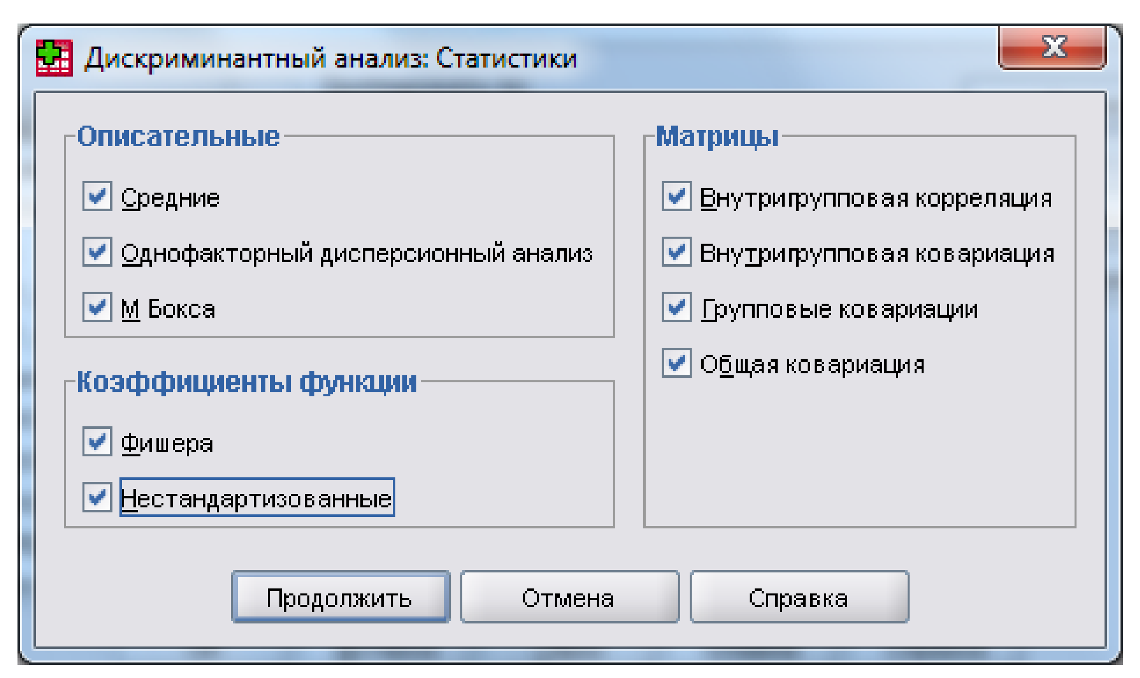



Рисунок 1.9 – Диалоговое окно «Дискриминантный анализ: Статистики»

Среди ограничений, касающихся статистических свойств дискриминантных переменных, есть равенство ковариационных матриц для различных классов. Это предположение проверили с помощью М критерия Бокса [12]:

$$M = (n - g)\log|A| - \sum_{k=1}^{g}(n_k - 1)\log|R|,$$

где $A$ - объединенная внутригрупповая матрица; $R$ - ковариационная матрица для *k-й* группы; $g$ – число групп; $n$ – общее число наблюдений по всем группам; $n_k$ – число наблюдений в $k$-й группе. Алгоритм получения данных матриц приведен ниже. Результат проверки равенства ковариационных матриц для различных классов с помощью М критерия Бокса приведен на рис. 1.10

Если значимость превышает 0,05, данные характеризуются многомерной нормальностью. Если значимость равна или меньше 0,05,

многомерная нормальность считается нарушенной, однако это не означает, что дискриминантный анализ не применим к данным; более того, включение в уравнение переменной, чье распределение отличается от нормального, иногда может улучшить результаты анализа. Отсутствие многомерной нормальности указывает лишь на то, что не все используемые переменные имеют нормальное распределение [13].

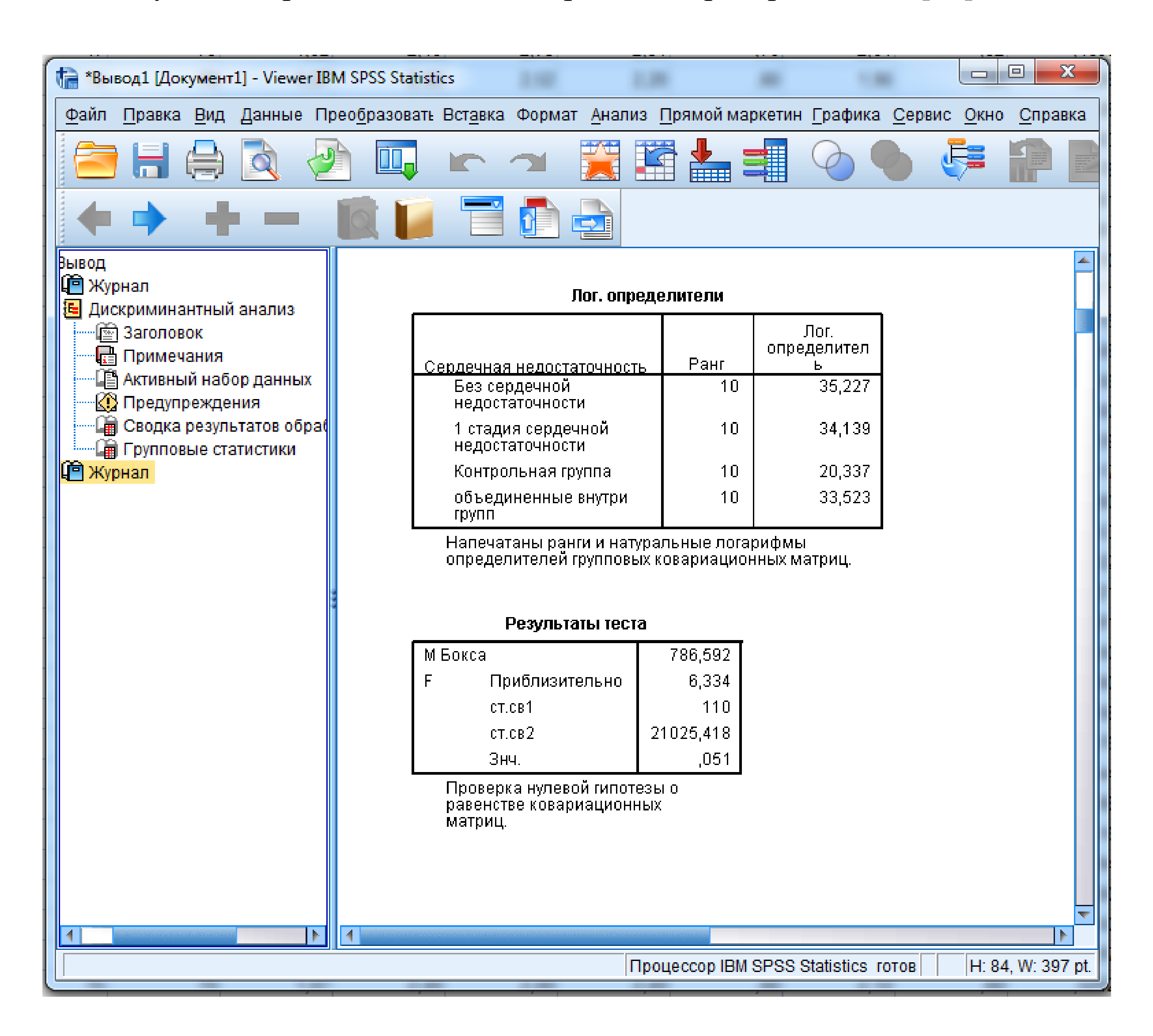



Рисунок 1.10 – Результаты теста Бокса

**Этап 5**. Определение коэффициентов дискриминантных функций.

Ядром дискриминантного анализа является построение так называемой дискриминантной функции:

$$DF_{km} = b_1 \cdot x_{1km} + b_2 \cdot x_{2km} + ... + b_p \cdot x_{pkm} + b_o, \qquad (1.1)$$

где $DF_{km}$ – значение дискриминантной функции для $m$-го пациента в $k$-й группе; $x_{ikm}$ – значение дискриминантной переменной $i$ для $m$-го пациента в $k$-й группе; $p$ – число дискриминантных переменных, $b_i$ – коэффициенты, которые предстоит оценить с помощью дискриминантного анализа.

Целью являются определение таких коэффициентов, чтобы по значениям дискриминантной функции можно было с максимальной чёткостью провести разделение на группы.

В дискриминантном анализе коэффициенты (или веса) $b_0$, $b_1$, $b_2$, ..., $b_p$ определяют таким образом, чтобы группы максимально отличались значениями дискриминантной функции. Это происходит тогда, когда отношение межгрупповой суммы квадратов к внутригрупповой сумме квадратов для дискриминантных показателей максимально.

Мы выбрали пошаговый (stepwise) дискриминантный анализ, при котором переменные вводятся последовательно, исходя из их способности различать (дискриминировать) группы. При пошаговом анализе «с включением» на каждом шаге просматриваются все переменные, и находится та из них, которая вносит наибольший вклад в различие между совокупностями. Эта переменная должна быть включена в модель на данном шаге, и происходит переход к следующему шагу.

Щелчком на кнопке «Метод» в окне «Дискриминантный анализ» (рис. 1.8) мы вошли в диалоговое окно «Дискриминантный анализ: Шаговый отбор», показанное на рис 1.11. Кнопка «Метод» доступна лишь в том случае, когда установлен переключатель «Шаговый отбор».

В этом окне при помощи группы переключателей «Метод» можно выбирать один из пошаговых методов составления дискриминантного уравнения, в области «Критерии» задать критерии для включения в дискриминантное уравнение и исключения из него предикторов, а также включить в окно вывода желаемые величины при помощи флажков в группе «Вывести на дисплей». Метод Лямбда Уилкса основан на минимизации коэффициента Уилкса после включения в уравнение регрессии каждого нового предиктора.

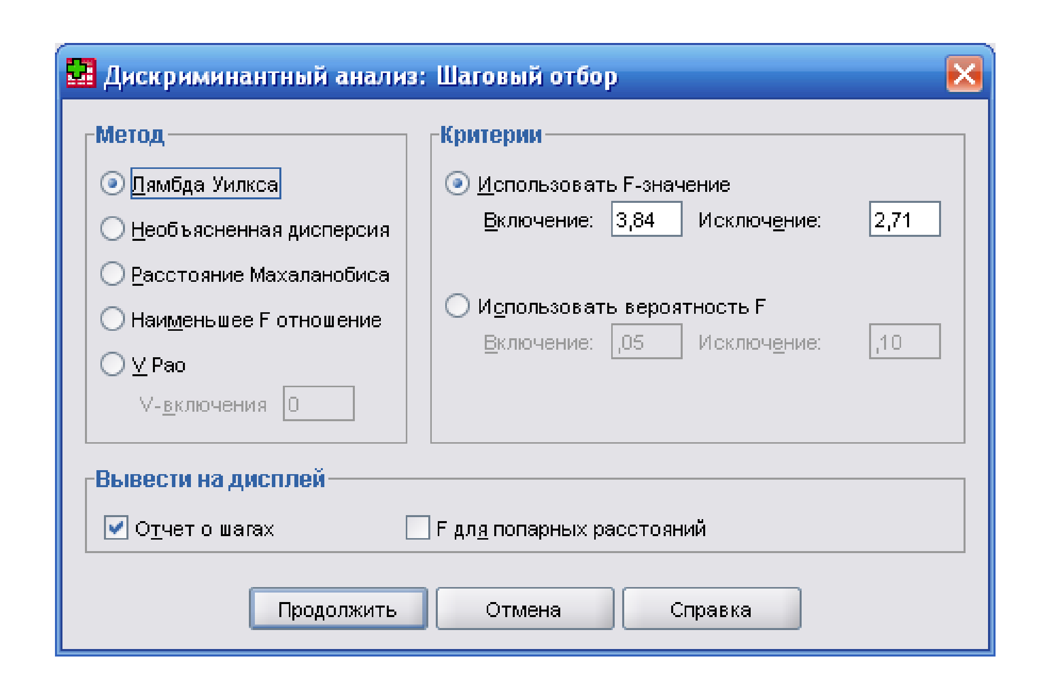



Рисунок 1.11 – «Дискриминантный анализ: Шаговый отбор»

По умолчанию критериями для включения и исключения предикторов являются пороговые значения F-критерия. Как правило F=3,84 соответствует величина уровня значимости p, приблизительно равная 0,05, а значению F=2,71 – величина p, приблизительно равная 0,1 [14].

Рассмотрим основные этапы получения коэффициентов «канонической дискриминантной функции.

Для получения коэффициентов $b_i$ канонической дискриминантной функции нужен статистический критерий различения групп. Проверка качества дискриминации (различия) основана на сравнении средних дискриминантной функции для исследуемых групп. Эти средние играют настолько важную роль в дискриминантном анализе, что получили свое название – центроиды (*centroids*). Центроидов столько, сколько групп, т.е. один центроид для каждой группы. Очевидно, что классификация переменных будет осуществляться тем лучше, чем меньше рассеяние точек относительно центроида внутри группы и чем больше расстояние между центроидами групп. Разумеется, что большая внутригрупповая вариация нежелательна, так как в этом случае любое заданное расстояние между двумя средними тем менее значимо в статистическом смысле, чем больше вариация распределений, соответствующих этим средним. Один из методов поиска наилучшей дискриминации данных заключается в

нахождении такой канонической дискриминантной функции DF, которая бы максимизировала отношение межгрупповой вариации к внутригрупповой:

$$\lambda = S / E, \qquad (1.2)$$

где $S$ – межгрупповая и $E$ – внутригрупповая матрицы рассеяния наблюдаемых переменных от средних.

Рассмотрим максимизацию отношения (1.2) для произвольного числа классов. Введем следующие обозначения:

$\bar{x}_i$ – среднее значение переменной $i$ по всем классам (общее среднее);

$x_{ikm}$ – значение переменной $i$ для *m-го* наблюдения в *k-й* группе;

$\bar{x}_{ik}$ – среднее значение переменной $i$ в *k-й* группе.

Рассмотрим задачу максимизации отношения (1.2) когда имеются g групп. Оценим сначала информацию, характеризующую степень различия между объектами по всему пространству точек, определяемому переменными групп. Для этого вычислим матрицу рассеяния $O$, которая равна сумме квадратов отклонений и попарных произведений наблюдений от общих средних $\bar{x}_i, i = 1,..., p$ по каждой переменной. Элементы матрицы $O$ определяются выражением:

$$o_{ij} = \sum_{k=1}^{g} \sum_{m=1}^{n_k} (x_{ikm} - \bar{x}_i)(x_{jkm} - \bar{x}_j).$$

Запишем это выражение в матричной форме. Обозначим $p$-мерную случайную векторную переменную $k$-й группы следующим образом:

$$X_k = \{x_{ikm}\}, i = 1,..., p, k = 1,...g, m = 1,...n_k.$$

Тогда объединенная $p$-мерная случайная векторная переменная всех групп будет иметь вид:

$$X = \left[ X_1 X_2 \ldots X_g \right].$$

Общее среднее этой $p$-мерной случайной векторной переменной будет равно вектору средних отдельных признаков:

$$\bar{x} = \left[ \overline{x_1} \overline{x_2} \ldots \overline{x_g} \right].$$

Матрица рассеяния от среднего при этом запишется в виде

$$O = \sum_{k=1}^{g} (X_k - \bar{x})(X_k - \bar{x})^T .$$

Если использовать векторную переменную объединенных переменных $X$, то матрица $O$ определится по формуле:

$$O = (X - \bar{x})(X - \bar{x})^T .$$

Матрица $O$ содержит полную информацию о распределении точек по пространству переменных. Диагональные элементы представляют собой сумму квадратов отклонений от общего среднего и показывают, как ведут себя наблюдения по отдельно взятой переменной. Внедиагональные элементы равны сумме произведений отклонений по одной переменной на отклонения по другой.

Если разделить матрицу $O$ на *(n - 1)*, то получим общую ковариационную матрицу $H$ (рис. А.1, Приложение А). В пакете SPSS общая ковариационная матрица получается путем установки опции «Общая ковариация» в диалоговом окне «Дискриминантный анализ: Статистики» (рис. 1.9). Для проверки условия линейной независимости переменных полезно рассмотреть вместо $O$ нормированную корреляционную матрицу.

Для измерения степени разброса объектов внутри групп рассмотрим матрицу $E$, которая отличается от $O$ только тем, что ее элементы определяются векторами средних для отдельных групп, а не вектором

средних для общих данных. Элементы внутригруппового рассеяния определятся выражением:

$$e_{ij} = \sum_{k=1}^{g} \sum_{m=1}^{n_k} (x_{ikm} - \bar{x}_{ik})(x_{jkm} - \bar{x}_{jk}) \tag{1.3}$$

Выражения в скобках являются отклонениями значений переменных от общего среднего. Если $i = j$, то сомножители равны, и получается среднеквадратичное отклонение. Таким образом, диагональные элементы представляют собой сумму квадратов отклонений от общего среднего. Они показывают как ведут себя наблюдения по отдельной переменной. При $i \neq j$ получаем сумму произведений отклонения по одной переменной на отклонение по другой. В этом состоит один из способов измерения корреляций (ковариаций) между двумя переменными, так как он показывает, насколько хорошо большое отклонение по одной переменной согласуется с большим отклонением по другой. Рассматривая целиком всю матрицу, мы имеем полную информацию о распределении точек по пространству, определяемому переменными.

Запишем выражение (1.3) в матричной форме. Данным групп будут соответствовать векторы средних:

$$\overline{x_k} = \left[ \overline{x_{1k}}\, \overline{x_{2k}} \ldots \overline{x_{pk}} \right].$$

Тогда матрица внутригрупповых вариаций запишется в виде:

$$E = \sum_{k=1}^{g} (X_k - \overline{x_k})(X_k - \overline{x_k})^T$$

Если разделить каждый элемент матрицы $E$ на $(n - g)$, то получится внутригрупповая ковариационная матрица $A$, она является взвешенным средним ковариационных матриц отдельных классов. Данная матрица $A$ (рис. А.2, Приложение А) была получена путем установки опции

«Внутригрупповая ковариация» в диалоговом окне «Дискриминантный анализ: Статистики» (рис. 1.9).

Степень зависимости двух переменных можно выяснить, исследуя их корреляцию. Для этого мы в окне «Дискриминантный анализ: Статистики» (рис. 1.9) отметили флажком «Внутригрупповая корреляция».

Коэффициент корреляции принимает значения от -1 до +1. Преобразование матрицы $E$ в матрицу коэффициентов корреляции (объединенную внутригрупповую матрицу) осуществляется делением каждого элемента $e_{ij}$ на квадратный корень произведения двух соответствующих диагональных элементов, а именно на $\sqrt{\sum_{k=1}^{g}\sum_{m=1}^{n_k}(x_{ikm}-\overline{x}_{ik})^2\sum_{k=1}^{g}\sum_{m=1}^{n_k}(x_{jkm}-\overline{x}_{jk})^2}$. На рис. А.3 (Приложение А) представлена объединенная внутригрупповая матрица *Q*, состоящая из соответствующих средних значений корреляционных матриц для уровней зависимой переменной.

Для получения групповой ковариационной матрицы сначала определяются элементы $g_{ij}$ :

$$g_{ij}=\sum_{m=1}^{n_k}(x_{ikm}-\overline{x}_{ik})(x_{jkm}-\overline{x}_{jk}).$$

Если элементы матрицы $g_{ij}$ разделить на $(n_k - 1)$, то получится групповая ковариационная матрица *R*. На рис. А.4 (Приложение А) представлена полученная в SPSS групповая ковариационная матрица *R*, представляющая собой отдельные ковариационные матрицы для каждого уровня зависимой переменной. Для получения данной матрицы в окне «Дискриминантный анализ: Статистики» (рис. 1.9) предварительно отметили флажком «Групповые ковариации».

Когда центроиды различных групп совпадают, то элементы матриц *O* и *E* будут равны. Если же центроиды групп различные, то разница $S=O–E$ будет определять межгрупповую сумму квадратов отклонений и попарных произведений. Если расположение групп в пространстве различается (т.е. их центроиды не совпадают), то степень разброса наблюдений внутри групп будет меньше межгруппового разброса.

Отметим, что элементы матрицы $S$ можно вычислить и по данным средних:

$$s_{ij} = \sum_{k=1}^{g} \sum_{m=1}^{n_k} (\overline{x_{ik}} - \overline{x}_i)(\overline{x_{jk}} - \overline{x}_j), \ i, j = 1, ..., p.$$

Матрицы $E$ и $S$ содержат всю основную информацию о зависимости внутри групп и между группами. Для лучшего разделения наблюдений на группы нужно подобрать коэффициенты дискриминантной функции из условия максимизации отношения межгрупповой матрицы рассеяния к внутригрупповой матрице рассеяния при условии ортогональности дискриминантных плоскостей. Тогда нахождение коэффициентов дискриминантных функций сводится к решению задачи о собственных значениях и векторах. Это утверждение можно сформулировать так: если спроектировать $g$ групп $p$-мерных выборок на ($g$ -1) пространство, порожденное собственными векторами $(v_{1k}, ..., v_{pk}), k = 1, ..., g-1,$ то отношение (1.2) будет максимальным, т. е. рассеивание между группами будет максимальным при заданном внутригрупповом рассеивании. Если бы мы захотели спроектировать $g$ выборок на прямую, при условии максимизации наибольшего рассеивания между группами, то следовало бы использовать собственный вектор $(v_{11}, ..., v_{p1}),$ соответствующий максимальному собственному числу $\lambda_1$. При этом дискриминантные функции можно получать: по нестандартизованным и стандартизованным коэффициентам.

Пусть $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$ и $v_{1k}, ..., v_{pk}$ соответственно собственные значения и векторы. Тогда условие (1.2) в терминах собственных чисел и векторов запишется в виде:

$$\lambda_i = \frac{\sum_k s_{ik} v_i}{\sum_k e_{ik} v_i},$$

что влечет следующее равенство, представленное в матричной форме:

$$(S-\lambda_i E)v_i = 0, v_i^T E v_j = \delta_{ij}\ , \quad (1.4)$$

где $\delta_{ij}$ - символ Кронекера - функция двух целых переменных, которая равна 1, если они равны, и 0 в противном случае [15]. Таким образом, решение уравнения $|S-\lambda E|=0$ позволяет нам определить компоненты собственных векторов, соответствующих дискриминантным функциям. Решение системы уравнений (1.4) можно получить путем использования разложения Холецкого матрицы $E^{-1}$ и решения задачи о собственных значениях [16 – 17].

Каждое решение, которое имеет свое собственное значение $\lambda_i$ и собственный вектор $v_i$, соответствует одной дискриминантной функции. Компоненты собственного вектора $v_i$ можно использовать в качестве коэффициентов дискриминантной функции. Однако при таком подходе начало координат не будет совпадать с главным центроидом. Для того чтобы начало координат совпало с главным центроидом, нужно нормировать компоненты собственного вектора:

$$b_{ik} = v_{ik}\sqrt{n-g},$$
$$b_{ok} = -\sum_{i=1}^{p} b_{ik}\overline{x_i}. \quad (1.5)$$

Нормированные коэффициенты (1.5) получены по нестандартизованным исходным данным, поэтому они называются нестандартизованными. Нормированные коэффициенты приводят к таким дискриминантным значениям, единицей измерения которых является стандартное квадратичное отклонение. При таком подходе каждая ось в преобразованном пространстве сжимается или растягивается таким образом, что соответствующее дискриминантное значение для данного объекта представляет собой число стандартных отклонений точки от главного центроида.

Расчет методом дискриминантных функций значений диагностических коэффициентов позволил выделить десять значимых для

определения хронической сердечной недостаточности показателей со своими нестандартизованными коэффициентами (табл. 1.1).

Таблица 1.1

Канонические коэффициенты дискриминантных функций для значимых показателей при определении состояния сердечно-сосудистой системы

| Показатели | Единицы измерения | Канонические коэффициенты, $b_i$ | |
|---|---|---|---|
| | | 1-я функция | 2-я функция |
| $x_1$ | пг/мл | - 0,019 | 0,080 |
| $x_2$ | пг/мл | - 0,016 | - 0,059 |
| $x_3$ | пг/мл | 0,001 | - 0,010 |
| $x_4$ | % клеток с маркером апоптоза | 0,053 | 0,130 |
| $x_5$ | нмоль/мл | - 0,003 | - 0,038 |
| $x_6$ | нг/мл/ч | 0,156 | 0,296 |
| $x_7$ | ед/мл | - 0,165 | 0,270 |
| $x_8$ | нмоль/сут | 0,001 | - 0,011 |
| $x_9$ | % | 0,120 | - 0,014 |
| $x_{10}$ | мл/м$^2$ | 0,020 | 0,009 |
| Константа | – | - 7,188 | - 1,122 |

Таким образом, развитие ХСН у подростков может быть описано следующими дискриминантными функциями [18 – 19]:

$$DF1 = -0{,}019*x_1 - 0{,}016*x_2 + 0{,}001*x_3 + 0{,}053*x_4 - 0{,}003*x_5 + +0{,}156*x_6 - 0{,}165*x_7 + 0{,}001*x_8 + 0{,}120*x_9 + 0{,}020*x_{10} - 7{,}188,$$

$DF2$ = 0,080* $x_1$ – 0,059* $x_2$ – 0,010* $x_3$ + 0,130* $x_4$ – 0,038* $x_5$ + +0,296* $x_6$ + 0,270* $x_7$ – 0,011* $x_8$ – 0,014* $x_9$ + 0,009* $x_{10}$ – 1,122,

где $x_1$ – интерлейкин-1 (пг/мл); $x_2$ – интерлейкин-6 (пг/мл); $x_3$ – фактор некроза опухоли ФНО-а (пг/мл); $x_4$ – CD-95 (% клеток, несущих маркер апоптоза); $x_5$ – циклический 3,5-аденозинмонофосфат (цАМФ, нмоль/мл); $x_6$ – ренин (нг/мл/час); $x_7$ – карбонилированные белки (Ед/мл); $x_8$ – норадреналин (нмоль/сут); $x_9$ – фракция выброса левого желудочка после физической нагрузки (%); $x_{10}$ – индекс конечного диастолического объема левого желудочка (мл/м2).

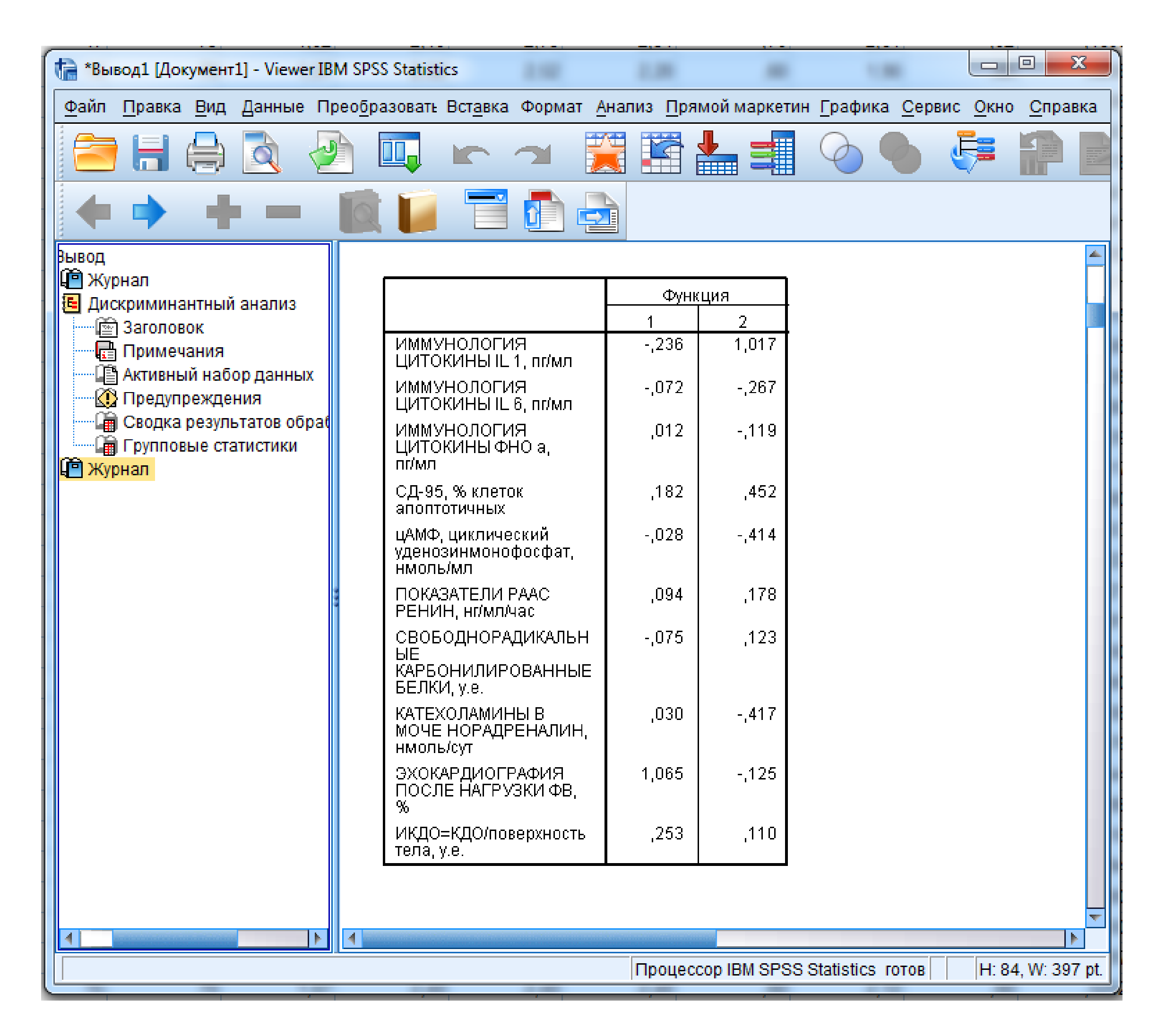
Стандартизованные коэффициенты можно получить преобразованием нестандартизованных коэффициентов к стандартизованной форме [20]:

$$c_{ik} = b_{ik}\sqrt{\frac{e_{ii}}{n-g}},$$

где $e_{ij}$ - сумма внутригрупповых квадратов $i$-й переменной, определяемой по формуле 3. Стандартизованные коэффициенты применяют для уменьшения размерности исходного признакового пространства переменных. Если абсолютная величина коэффициента для данной переменной для всех дискриминантных функций мала, то эту переменную можно исключить, тем самым сократив число переменных. Полученные стандартизованные коэффициенты канонической дискриминантной функции приведены на рис. 1.12

**Этап 6.** Определение коэффициентов канонической корреляции.

Оценка степени удачного распределения на группы, полезность дискриминантных функций и количество функций, имеющих реальное содержание при определении различий между группами, оценены с помощью коэффициентов канонической корреляции (табл. 1.2).




| | Функция | |
|---|---|---|
| | 1 | 2 |
| ИММУНОЛОГИЯ ЦИТОКИНЫ IL 1, пг/мл | -,236 | 1,017 |
| ИММУНОЛОГИЯ ЦИТОКИНЫ IL 6, пг/мл | -,072 | -,267 |
| ИММУНОЛОГИЯ ЦИТОКИНЫ ФНО а, пг/мл | ,012 | -,119 |
| СД-95, % клеток апоптотичных | ,182 | ,452 |
| цАМФ, циклический уденозинмонофосфат, нмоль/мл | -,028 | -,414 |
| ПОКАЗАТЕЛИ РААС РЕНИН, нг/мл/час | ,094 | ,178 |
| СВОБОДНОРАДИКАЛЬНЫЕ КАРБОНИЛИРОВАННЫЕ БЕЛКИ, у.е. | -,075 | ,123 |
| КАТЕХОЛАМИНЫ В МОЧЕ НОРАДРЕНАЛИН, нмоль/сут | ,030 | -,417 |
| ЭХОКАРДИОГРАФИЯ ПОСЛЕ НАГРУЗКИ ФВ, % | 1,065 | -,125 |
| ИКДО=КДО/поверхность тела, у.е. | ,253 | ,110 |

Рисунок 1.12 – Стандартизованные коэффициенты канонической дискриминантной функции

Полученные собственные значения приведены в порядке убывания их величин. Величина собственного значения связана с дискриминирующими возможностями функции: чем больше собственное значение, тем лучше различие. Таким образом, первая функция обладает наибольшими дискриминирующими возможностями, вторая – обеспечивает максимальное различие после первой. Фактические числа, представляющие собственные значения, трудно интерпретировать непосредственно.

Чтобы облегчить сравнение, собственным значением приписывают процент дисперсии. Для этого сначала суммируют все собственные значения (13,010+1,264=14,274), затем делят каждое собственное значение на общую сумму (13,010/14,274=0,911 и 1,264/14,274=0,089). Так первая

функция содержит 91,1% общих дискриминантных возможностей, а вторая - 8,9%.

Таблица 1.2

Характеристика дискриминантных возможностей функций

| Функция | Собственное значение $\lambda_i$ | % дисперсии | Общий % | Каноническая корреляция $r_i$ |
|---|---|---|---|---|
| 1 | 13,010 | 91,1 | 91,1 | 0,964 |
| 2 | 1,264 | 8,9 | 100,0 | 0,747 |

Мы видим, что вторая функция малозначима, но не знаем, можно ли ею пренебречь, так как нет правила, которое помогло бы определить, как велико должно быть относительное процентное содержание, чтобы функция представляла для нас интерес. Относительное процентное содержание показывает только то, насколько функция слабее по сравнению с другими и ничего не добавляет к определению различий между группами.

Для того чтобы определить, представляет ли интерес, например, вторая функция необходимо определить коэффициент канонической корреляции.

В дискриминантном анализе различают две задачи: интерпретации и классификации. Задача интерпретации – определение количества, значимости дискриминантных функций и их значений для объяснения различий между классами. Задача классификации – определение класса, к которому принадлежит новый объект. В некоторых случаях нет необходимости решать задачу интерпретации. В этой ситуации используют простые классифицирующие функции, называемые дискриминантными, а при решении задачи интерпретации – канонические дискриминантные функции. Каноническая корреляция показывает степень зависимости между дискриминантными функциями и классами. Нулевое значение говорит об отсутствии связи, а большие числа (всегда положительные) означают большую степень зависимости (максимальное значение равно 1).

Следовательно, оценку реальной полезности дискриминантной функции получили, определив коэффициент канонической корреляции $r_i$ [21]:

$$r_i = \sqrt{\frac{\lambda_i}{1+\lambda_i}}, \qquad (1.6)$$

где $i$ — номер соответствующей дискриминантной функции;

$\lambda_i$ – собственное значение.

Например, коэффициент канонической корреляции для первой дискриминантной функции определяется как:

$$r_1 = \sqrt{\frac{13,010}{1+13,010}} = 0,964.$$

Значение коэффициента канонической корреляции, первой функции (0,964) позволяет сделать вывод о существовании достаточно высокой связи между ХСН и значением первой дискриминантной функции, что и предполагалось процентом дисперсии этой функции. Также выявлена высокая связь между случаями развития ХСН (0,747) и значение второй дискриминантной функции. Качественная оценка силы связи $r_{XY}$ величин $X$ и $Y$ была сделана на основании шкалы Чеддока (табл. 1.3) [22].

Таблица 1.3

Таблица анализа силы связи между переменными

| Значение $r_i$ | Интерпретация |
|---|---|
| от 0 до 0,3 | Очень слабая |
| от 0,3 до 0,5 | Слабая |
| от 0, 5 до 0,7 | Средняя |
| от 0,7 до 0, 9 | Высокая |
| от 0,9 до 1 | Очень высокая |

**Этап 7**. Определение значимости дискриминантных функций методом λ-статистики Уилкса.

Оценка значимости дискриминантных функций была проверена методом λ-статистики Уилкса (табл. 1.4).

λ-статистика Уилкса – это мера различий между классами по нескольким дискриминантным переменным. λ-статистика Уилкса принимает значения, обратные привычным, и оценивает остаточную дискриминантную способность Чем ближе значение λ к 0, тем лучше различие классов, а чем ближе к 1, тем различие хуже (классы совпадают).

Определяется λ-статистика Уилкса согласно формуле:

$$\lambda_i^* = \prod_{i=k+1}^{g} \frac{1}{1+\lambda_i}, \tag{1.7}$$

где $k$ – число вычисленных функций;

$\lambda_i$ – собственное значение.

Таблица 1.4

Определение значимости дискриминантных функций методом λ-статистики Уилкса

| Тест функций | Лямбда Уилкса $\lambda_i^*$ | $\chi_i^2$ | Степени свободы, $df_i{}^*$ | Значимость $P_i$ |
|---|---|---|---|---|
| 1 | 0,032 | 523,746 | 20 | 0,0001 |
| 2 | 0,442 | 123,819 | 9 | 0,0001 |

Например, λ-статистика Уилкса для первой функции определяется как:

$$\lambda_1^* = \frac{1}{1+13,010} \cdot \frac{1}{1+1,264} = 0,032\,.$$

Для второй дискриминантной функции λ-статистика Уилкса рассчитывается как:

$$\lambda_2^* = \frac{1}{1+1,264} = 0,442\,.$$

Поскольку $\lambda$ является «обратной» мерой, этот результат означает, что все используемые переменные эффективно участвуют в различении классов.

Из табл. 1.4 видно, что вычисление первой функции уменьшает количество оставшейся информации, и величина $\lambda_i^*$ становится равной 0,442 (остается небольшой).

**Этап 8.** Расчет критериального значения $\chi_i^2$, которое сравнивается с критическим, с заданным числом степеней свободы $df_i^* = (p-(k-1))(g-(k-1)-1$.

Число степеней свободы зависит от количества дискриминантных переменных $p$ и количества классов $g$:

$$\chi_i^2 = -\left[n - \left(\frac{p+g}{2}\right) - 1\right] \ln \lambda_i^*, \tag{1.8}$$

Так, например, число степеней свободы для первой и второй дискриминантных функции определяется как:

$$df_1^* = (10-(1-1))(3-(1-1)-1)=20,$$

$$df_2^* = (10-(2-1))(3-(2-1)-1)=9.$$

Критериальное значение $\chi_i^2$, например, для второй дискриминантной функции определяется как:

$$\chi_2^2 = -\left[159 - \left(\frac{10+3}{2}\right) - 1\right] \ln 0,442 \approx 123,819.$$

В нашем случае, уровень значимости $P_i<0,0001$, то есть различие между группами значимое.

**Этап 10.** Расчет квадрата расстояния Махалонобиса до среднего значения группы значений дискриминантной функции.

Рассчитанные значения дискриминантных функций рассматриваются как точки в некотором пространстве. Для каждой группы можно рассчитать центр группирования (центроид) [23]:

$$df_k = \frac{\sum_{m=1}^{n_k} DF_{km}}{n_k}.$$

В табл. 5 приведены полученные в SPSS значения центроидов для каждой группы.

Таблица 1.5

Значения центроидов для каждой группы

| Группы | Функция | |
|---|---|---|
| | 1 | 2 |
| Без сердечной недостаточности | 2,099 | 2,374 |
| 1 стадия сердечной недостаточности | 3,409 | -,815 |
| Контрольная группа | -4,153 | -,166 |

Для каждого объекта рассчитывают расстояние от него до каждой точки группирования. Для этого используют квадрат расстояния Махалонобиса (расстояние от объекта, который необходимо классифицировать, до центра класса) [24].

Для того чтобы в SPSS получить информацию о квадрате расстояния Махалонобиса в диалоговом окне «Дискриминантный анализ». Выбор метода (рис. 1.6) выбрали кнопку «Классифицировать» и активировали опцию «Поточечные результаты» (рис. 1.13).

В полученной в SPSS после включения опции «Поточечные результаты» (рис. 1.13) таблице А.2 (Приложение А) в столбце «Квадрат расстояния Махалонобиса до центра» приводится квадрат расстояния Махалонобиса до среднего значения группы значений дискриминантной функции. В столбце «Значения дискриминантных функций» приводятся величины, получаемые при подстановке значений переменных пациента в уравнение дискриминантных функций.

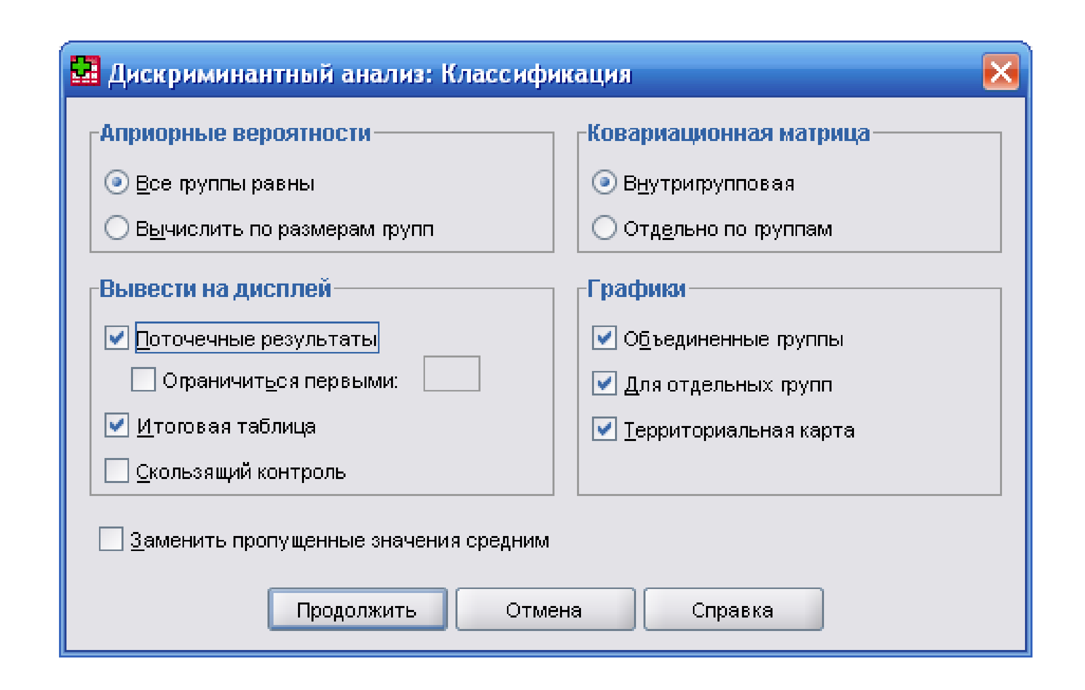



Рисунок 1.13 – Диалоговое окно «Дискриминантный анализ: Классификация»

**Этап 11**. Определение группы, к которой принадлежит пациент.

Относя объект к ближайшему классу в соответствии со значением $D^2$, мы неявно приписываем его к тому классу, для которого он имеет наибольшую вероятность принадлежности.

Группа, к которой фактически принадлежит пациент, отображается в колонке «Фактическая группа» (табл. А.2, Приложение А). В следующих колонках содержится информация о прогнозе принадлежности к группе, сделанном на основании значений дискриминантных функций.

P(D>d | G=g) – это вероятность того, что пациент, принадлежащий к прогнозируемой группе, действительно имеет значения параметров, соответствующие дискриминантной функции или некоторые другие крайние значения.

P(G=g|D=d) – мера принадлежности к одной из двух групп. Это вероятность того, что некоторый случай принадлежит к прогнозированной группе, которая рассчитывается на основе подстановки в дискриминантную функцию значений набора переменных, соответствующих данному случаю. Вероятность того, что данный случай принадлежит к другой группе получается вычитанием меры принадлежности из 1.

**Этап 12.** Построение территориальной карты на основе значений дискриминантных функций.

Запрос на вывод территориальной карты осуществляется с помощью опции «Территориальная карта» в диалоговом окне «Дискриминантный анализ: Классификация» (рис. 1.13).

На основании значений дискриминантный функций была построена территориальная карта распределения больных в зависимости от диагноза (рис. 1.14), где 1 – группа, у которой отсутствует сердечная недостаточность; 2 – группа, у которой начальная стадия сердечной недостаточности; 3 – группа здоровых подростков; * – центр группы.

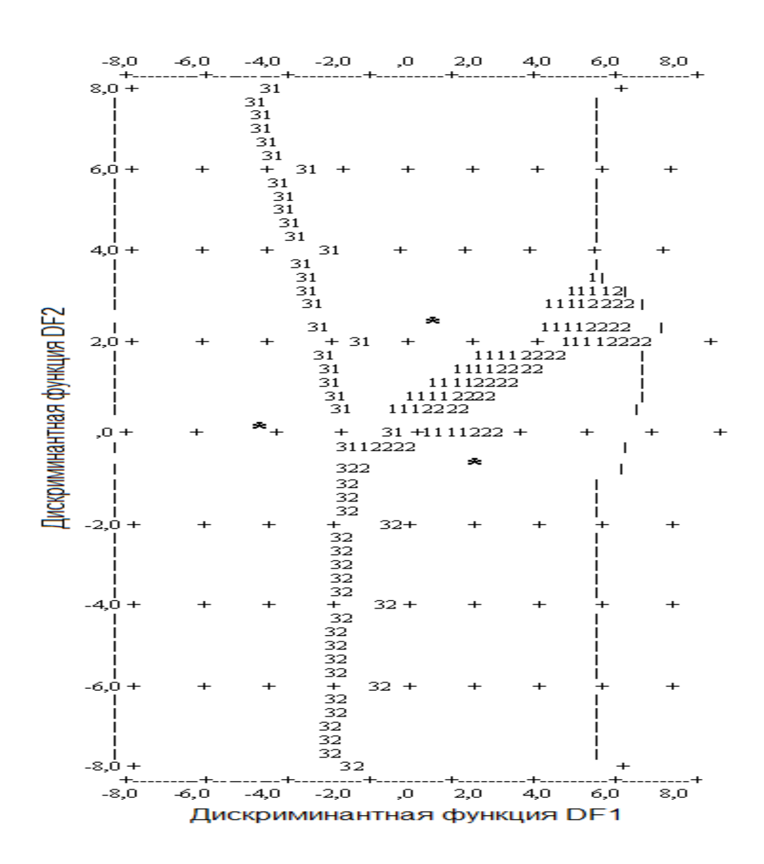



Рисунок 1.14 – Территориальная карта распределения больных с хронической сердечной недостаточностью и без неё

**Этап 13.** Построение графика распределения пациентов в зависимости от диагноза.

В диалоговом окне «Дискриминантный анализ: Классификация» (рис. 1.13) сделали запрос на вывод графика распределения, установив флажок в поле «Объединенные группы».

На рис. 1.15 представлен график распределения пациентов зависимости от диагноза, где по оси $X$ располагаются значения дискриминантной функции 1 (*DF1*), а по оси $Y$ – значения дискриминантной функции 2 (*DF2*).

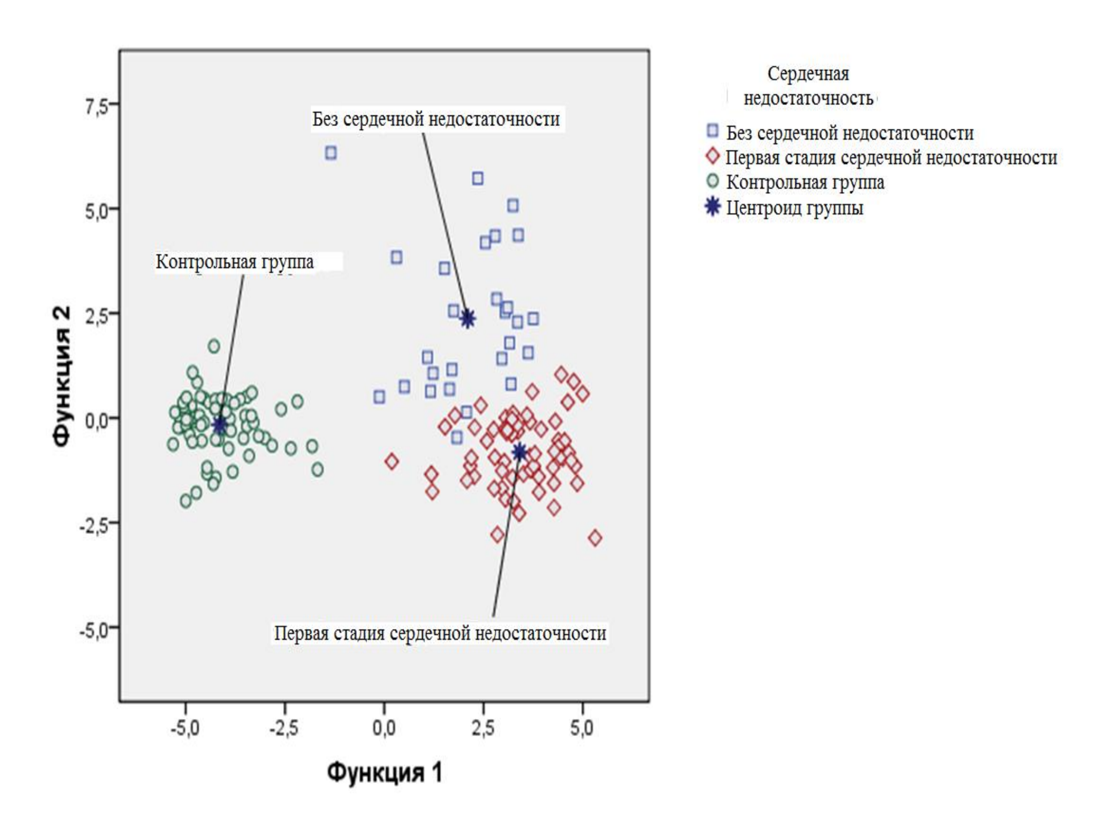



Рисунок 1.15 – График распределения пациентов с различными состояниями сердечно-сосудистой системы

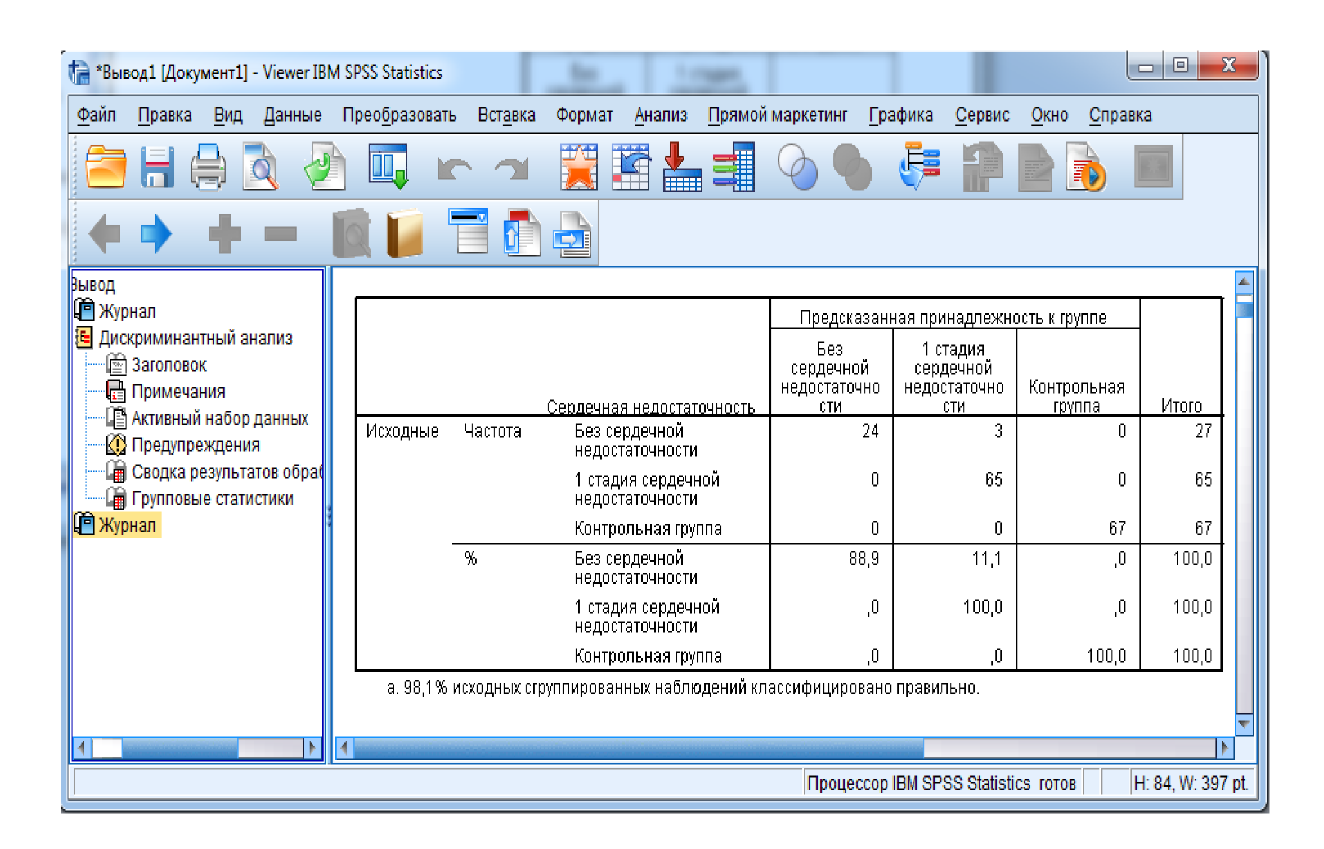
**Этап 14.** Определение $\tau$-статистики ошибок и доли правильной диагностики. Стандартизованной мерой эффективности для любого количества классов является $\tau$-статистика ошибок [25 – 27]:

$$\tau = \frac{n_c - \sum_{k=1}^{g} p_k n_k}{n - \sum_{k=1}^{g} p_k n_k}, \qquad (1.9)$$

где $n_c$ – число правильно классифицированных объектов, а $p_k$ – априорная вероятность принадлежности к классу.

Выражение $\sum_{k=1}^{g} p_k n_k$ представляет собой число объектов, которые будут правильно предсказаны при случайной классификации их по группам пропорционально априорным вероятностям. Если все группы считаются равноправными, то априорные вероятности полагаются равными единице, деленной на число классов. Максимальное значение $\tau$-статистики равно 1 и оно достигается в случае безошибочного предсказания. Нулевое значение указывает на неэффективность процедуры, $\tau$-статистика может принимать и отрицательные значения, что свидетельствует о плохом различении или вырожденном случае. Поскольку $n_c$ должно быть целым числом, числитель может стать отрицательным чисто случайно, когда нет различий между группами.

Вывод в SPSS таблицы, представленной на рис 1.16, осуществляется путем установки опции «Итоговая таблица» в диалоговом окне «Дискриминантный анализ: Классификация» (рис. 1.14).




| | | | Предсказанная принадлежность к группе | | | |
|---|---|---|---|---|---|---|
| | | Сердечная недостаточность | Без сердечной недостаточности | 1 стадия сердечной недостаточности | Контрольная группа | Итого |
| Исходные | Частота | Без сердечной недостаточности | 24 | 3 | 0 | 27 |
| | | 1 стадия сердечной недостаточности | 0 | 65 | 0 | 65 |
| | | Контрольная группа | 0 | 0 | 67 | 67 |
| | % | Без сердечной недостаточности | 88,9 | 11,1 | ,0 | 100,0 |
| | | 1 стадия сердечной недостаточности | ,0 | 100,0 | ,0 | 100,0 |
| | | Контрольная группа | ,0 | ,0 | 100,0 | 100,0 |

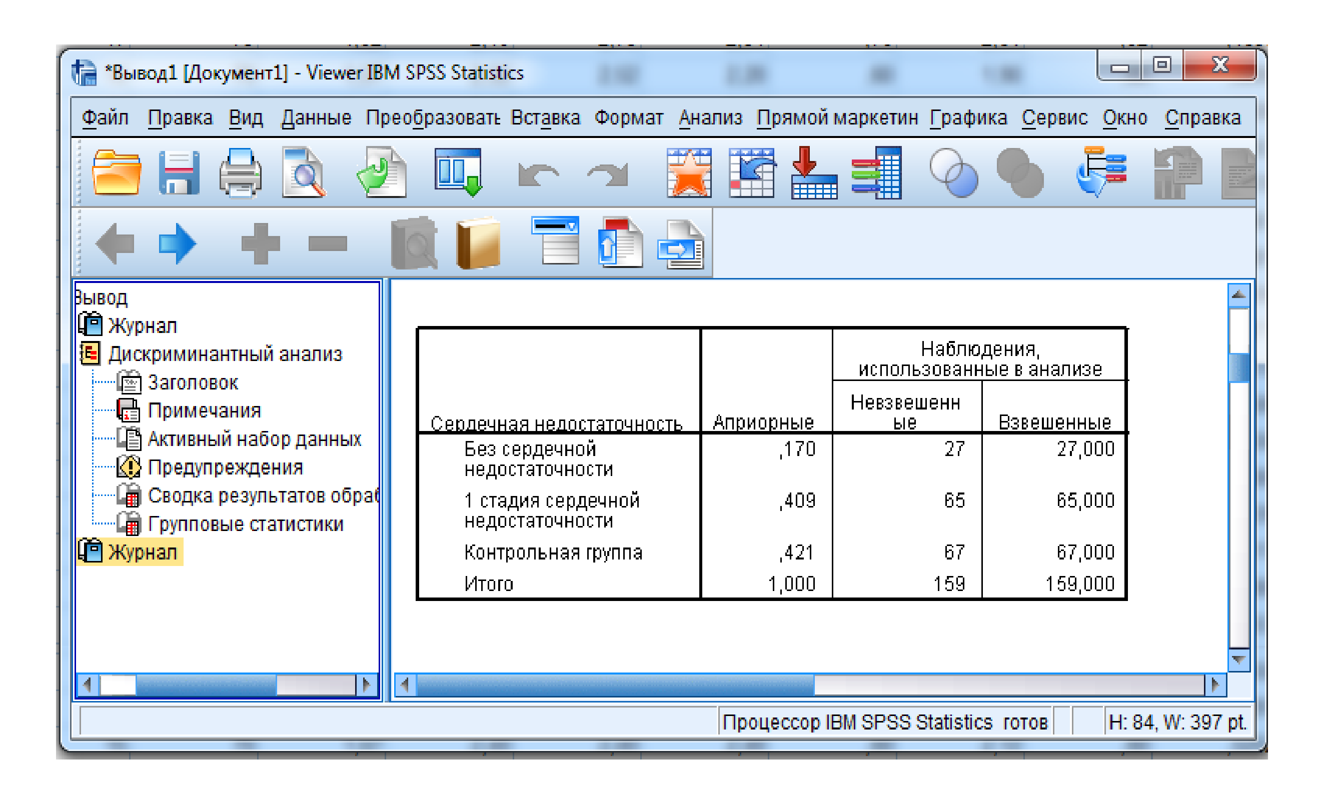
a. 98,1 % исходных сгруппированных наблюдений классифицировано правильно.

Рисунок 1.16 – Результаты классификации ХСН в SPSS

К группе 1 (подростки с заболеваниями сердца без ХСН) относится 27 наблюдений. 24 наблюдений (88, 9 %) были определены корректно, а 3 (11,1 %) по ошибке отнесены к группе 2. В группе 2 (подростки с начальной стадией ХСН) насчитывается 65 наблюдений. Все они были определены корректно. В группе 3 (контроль) насчитывается 67 наблюдения. Все они также были определены корректно (рис. 1.16).

Анализ результатов, полученных с помощью SPSS, показал, что прогноз для исследуемых групп сделан верно (корректно классифицированы в общей сложности 98,1% всех случав заболеваний сердца).

Для определения τ-статистики ошибок и определения априорных вероятностей $p_k$ для каждой группы пациентов установили в поле диалогового окна «Дискриминантный анализ: Классификация» (рис. 1.13) флажок «Вычислить по размерам групп» (рис. 1.17).




| Сердечная недостаточность | Априорные | Наблюдения, использованные в анализе: Невзвешенные | Взвешенные |
|---|---|---|---|
| Без сердечной недостаточности | ,170 | 27 | 27,000 |
| 1 стадия сердечной недостаточности | ,409 | 65 | 65,000 |
| Контрольная группа | ,421 | 67 | 67,000 |
| Итого | 1,000 | 159 | 159,000 |

Рисунок 1.17 – Априорные вероятности для всех групп

$$\sum_{k=1}^{g} p_k n_k = (0{,}170 \cdot 27) + (0{,}409 \cdot 65) + (0{,}421 \cdot 67) = 59{,}382.$$

$$\tau = \frac{156 - 59{,}382}{159 - 59{,}382} = 0{,}9698.$$

Т.е. классификация с помощью дискриминантных функций делает на 96,9% ошибок меньше, чем ожидалось при случайной классификации.

Нередко как показатель качества алгоритма диагностики (прогностической «силы») используют долю правильной диагностики:

$$\mu = p_1\kappa + p_2\lambda + p_3\eta, \qquad (1.10)$$

где κ – доля правильной диагностики в 1-й группе;

λ – доля правильной диагностики во 2-й группе;

η– доля правильной диагностики в 3-й группе;

$p_k$ (k=1, 2, 3; $p_1+p_2+p_3=1$) – доли групп в объединенной совокупности.

$p_1=27/159=0,170$; $p_2=65/159=0,409$; $p_3=67/159=0,421$.

$$\sum_{k=1}^{g} p_k n_k = 0,170 + 0,409 + 0,421 = 1.$$

$$\kappa = \frac{24}{27} = 0,889;$$

$$\lambda = \frac{65}{65} = 1;$$

$$\eta = \frac{67}{67} = 1.$$

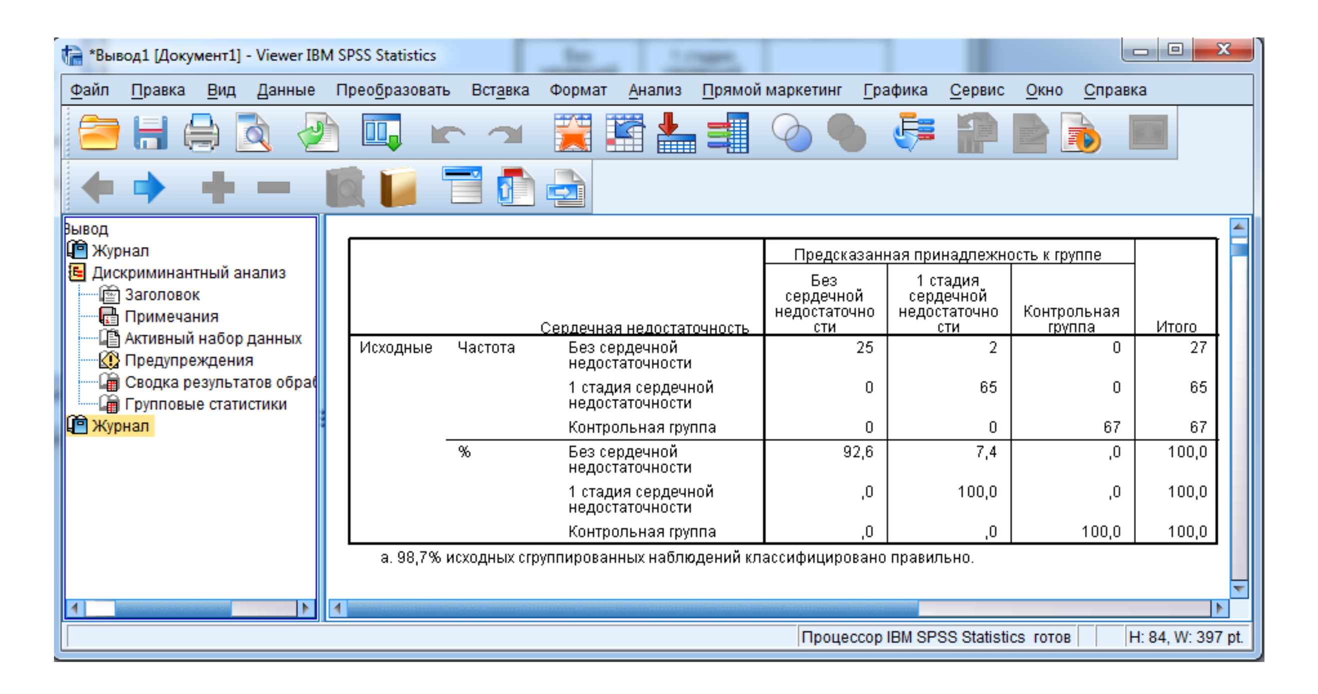
$$\mu = 0,170 \cdot 0,889 + 0,409 \cdot 1 + 0,421 \cdot 1 = 0,9811.$$

Данный результат соответствует полученному в SPSS (рис. 1.16).

Для корректного применения дискриминантных функции при решении поставленной задачи использовались две выборки пациентов, по 159 человек каждая: одна для вычисления дискриминантной функции (анализируемая), вторая – проверочная (для проверки результатов расчета на основании первой выборки). Т.е была проведена кросс-проверка (перекрестная проверка).

В результате проведения клинической апробации были получены следующие результаты (рис. 1.18).

К группе 1 относится 27 наблюдений. 25 наблюдений (92,6 %) было определено корректно, 2 ошибочно отнесены к группе 2. В группе 2 насчитывается 65 наблюдений. 65 из них были определены корректно. В группе 3 насчитывается 67 наблюдений. Корректно определено 67.




| | | Сердечная недостаточность | Предсказанная принадлежность к группе: Без сердечной недостаточности | 1 стадия сердечной недостаточности | Контрольная группа | Итого |
|---|---|---|---|---|---|---|
| Исходные | Частота | Без сердечной недостаточности | 25 | 2 | 0 | 27 |
| | | 1 стадия сердечной недостаточности | 0 | 65 | 0 | 65 |
| | | Контрольная группа | 0 | 0 | 67 | 67 |
| | % | Без сердечной недостаточности | 92,6 | 7,4 | ,0 | 100,0 |
| | | 1 стадия сердечной недостаточности | ,0 | 100,0 | ,0 | 100,0 |
| | | Контрольная группа | ,0 | ,0 | 100,0 | 100,0 |

а. 98,7% исходных сгруппированных наблюдений классифицировано правильно.

Рисунок 1.18 – Результаты классификации

Анализ полученных результатов показал, что прогноз для исследуемых групп сделан верно (корректно классифицированы в общей сложности 98, 7 % всех случав заболеваний сердца).

Результаты оценки показателя качества алгоритма диагностики (прогностической «силы») приведены ниже.

$p_1$=27/159=0,170; $p_2$=65/159=0,409; $p_3$=67/159=0,421.

$$\sum_{k=1}^{g} p_k n_k \quad 0{,}170+0{,}409+0{,}421=1.$$

$$\kappa=\frac{25}{27}=0{,}926;$$

$$\lambda = \frac{65}{65} = 1;$$

$$\eta = \frac{67}{67} = 1.$$

$$\mu = 0,170 \cdot 0,926 + 0,409 \cdot 1 + 0,421 \cdot 1 = 0,9874.$$

**Этап 15**. Формирование диагностического заключения.

Таким образом, используя аппарат дискриминантного анализа, на основе измерения морфофункциональных характеристик сердца пациента, показателей систем нейрогуморальной регуляции, иммунновоспалительной активации и свободнорадикальных процессов классифицировали пациентов, т.е. определили группу, к которой принадлежит пациент: (с заболеваниями сердечно-сосудистой системы без признаков ХСН; с заболеваниями сердца и начальной стадией ХСН; контрольная группа, которую составили практически здоровые подростки. Это позволило предсказать группы (наличие сердечной недостаточности) для новых объектов из той же совокупности.

В данной главе показана возможность и целесообразность использования дискриминантного анализа для идентификации состояния больных ХСН.

# ГЛАВА 2. ОПРЕДЕЛЕНИЕ ДИСФУНКЦИИ МИОКАРДА У ПОДРОСТКОВ С ИСПОЛЬЗОВАНИЕМ ПАКЕТА ПРОГРАММ STATISTICA

Одним из ранних проявлений доклинической стадии ХСН является систолическая дисфункция миокарда. Это состояние у детей и подростков часто протекает бессимптомно, латентно и выявить его можно только с помощью проведения функциональных проб. Одной из таких проб является проба с дозированной физической нагрузкой, позволяющая определить оптимальный режим физических нагрузок для подростков с различным уровнем здоровья. В этом плане актуальна разработка математической модели прогнозирования функционального состояния миокарда у пациентов, которая позволит определить больных с адекватной реакцией сердечно-сосудистой системы на физическую нагрузку и с неадекватной, т.е. систолической дисфункцией.

**Этап 1.** Сбор информации.

Для определения систолической дисфункции миокарда у подростков была исследована совокупность объектов (137 пациентов 10 - 18 лет), разделенная на три группы: 1-я группа - пациенты с адекватной реакцией на физическую нагрузку (52 подростка); 2-я группа - пациенты с неадекватной реакцией на физическую нагрузку (23 подростка); 3-я группа - контрольная (62 здоровых сверстника).

**Этап 2.** Определение оптимального объема выборки.

Так как из предварительного исследования известно, что среди всех пациентов больные с позитивной динамикой составляют 37,9%, при уровне надежности $\gamma = 0{,}95$ и предельной ошибки выборки 5-6% необходимый объем выборки составляет:

$$n_{\omega} = \frac{t^2 \cdot \omega \cdot (1-\omega)}{\Delta_{\omega}^2} = 119 \div 168 \text{ человек,}$$

где $t$ - стандартизированный коэффициент Стьюдента.

**Этап 3.** Кодирование информации.

Для определения признаков, которые существенно влияют на процесс дифференциальной диагностики систолической дисфункции миокарда, было исследовано 83 признака, включающие в себя паспортные данные, данные анамнеза заболевания и жизни, клинические симптомы, показатели эхокардиографии (после нагрузки) и т.д.. Все признаки были закодированы и поставлены в соответствие 83-мерному вектору, учитывающему отсутствие, присутствие, направленность и величину каждого показателя. Кодировочная таблица представлена в табл. Б.1 [29].

Для ввода данных двойным щелчком левой клавиши мыши на значке STATISTICA 10 запустили программу STATISTICA. В открывшемся редакторе данных (рис. 2.1) ввели все подготовленные данные. Для этого дважды нажимали на ячейке с надписью «Var», после чего и появилось окно «Variable», в котором ввели всю информацию о переменных.

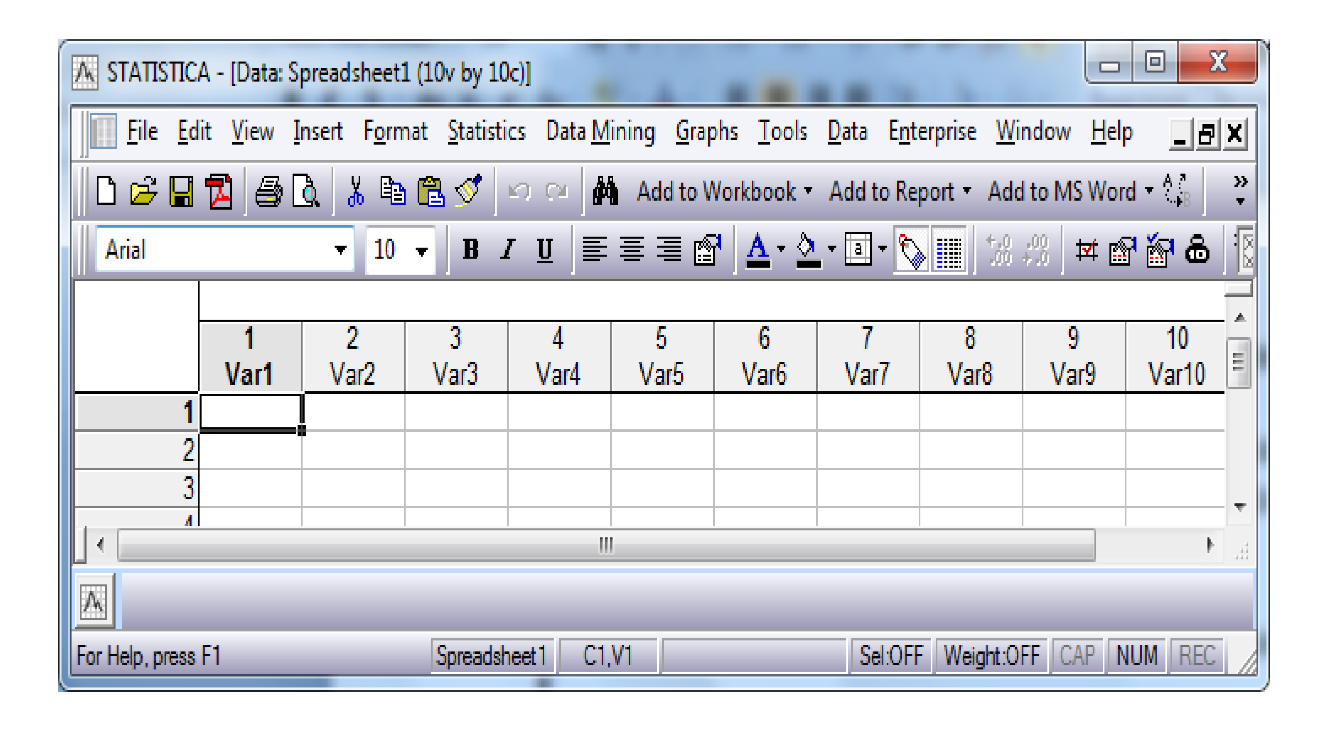



Рисунок 2.1 – Редактор данных: просмотр данных

Для кодирования качественных переменных, например, показателя raznica, по которому мы и будем группировать исследуемых пациентов, нажали на кнопку «Text Labels» диалогового окна редактирования переменных (рис. 2.2)., после чего открывается диалоговое окно «Text Labels Editor».

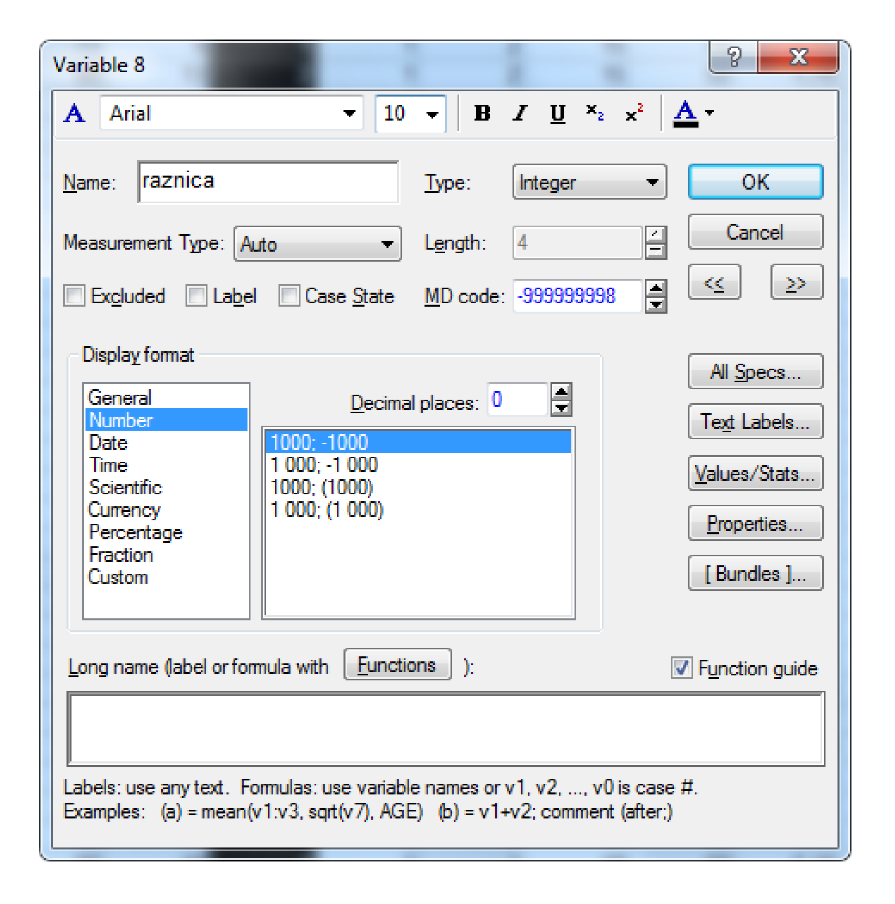



Рисунок 2.2 – Диалоговое окно «Variable»

Метки значений определялись следующим образом (рис. 2.3):

1) вначале ввели в поле «Text Label» название 1 группы «положительная нагрузка», а затем в поле «Numeric» число «1»;

2) повторили эти действия для значений «2» — «отрицательная нагрузка» и «3» — «норма».

Таким образом, были закодированы все качественные показатели.

Результат ввода всех значений в диалоговом окне показан на рис. 2.4.

**Этап 4.** Определение дискриминантных переменных, включаемых в модель.

Задача определения количества дискриминантных переменных решается путем процедуры последовательного отбора.

Нам необходимо принять решение по отбрасыванию переменных, которые заведомо не могут быть хорошими предикторами [30].

После ввода всех необходимых данных для проведения дискриминантного анализа выбрали в меню «Statistics» пункт «Multivariate

Exploratory Techniques → Discriminant Analysis» (рис. 2.5), после чего открывается диалоговое окно «Discriminant Function Analysis» (рис. 2.6).

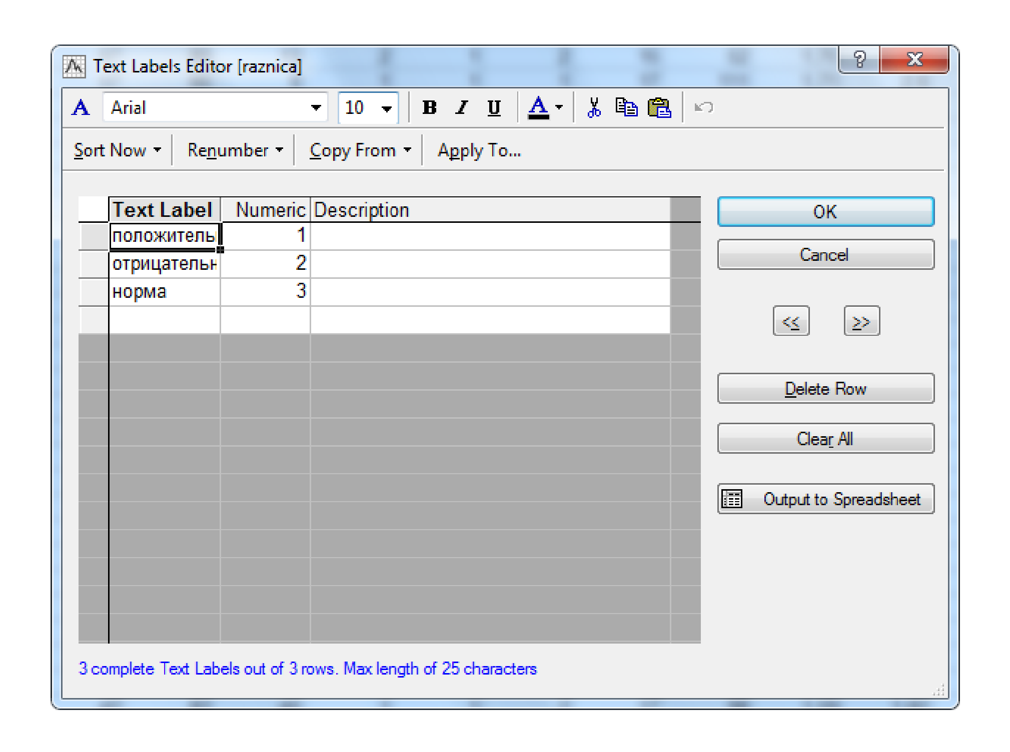
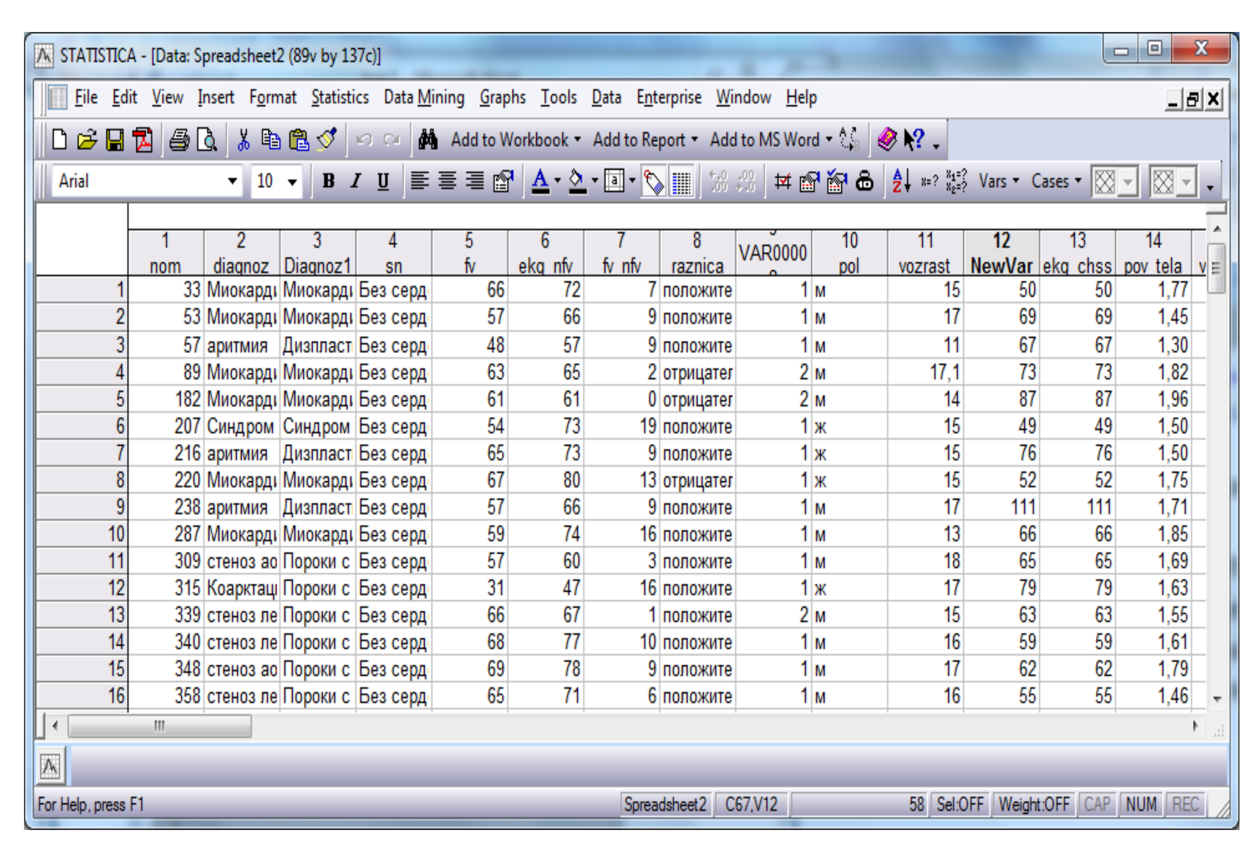



Рисунок 2.3 – Диалоговое окно «Text Labels Editor»




| | 1 nom | 2 diagnoz | 3 Diagnoz1 | 4 sn | 5 fv | 6 ekg_nfv | 7 fv_nfv | 8 raznica | VAR0000 | 10 pol | 11 vozrast | 12 NewVar | 13 ekg_chss | 14 pov_tela |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 33 | Миокарди | Миокарди | Без серд | 66 | 72 | 7 | положите | 1 | м | 15 | 50 | 50 | 1,77 |
| 2 | 53 | Миокарди | Миокарди | Без серд | 57 | 66 | 9 | положите | 1 | м | 17 | 69 | 69 | 1,45 |
| 3 | 57 | аритмия | Дизпласт | Без серд | 48 | 57 | 9 | положите | 1 | м | 11 | 67 | 67 | 1,30 |
| 4 | 89 | Миокарди | Миокарди | Без серд | 63 | 65 | 2 | отрицател | 2 | м | 17,1 | 73 | 73 | 1,82 |
| 5 | 182 | Миокарди | Миокарди | Без серд | 61 | 61 | 0 | отрицател | 2 | м | 14 | 87 | 87 | 1,96 |
| 6 | 207 | Синдром | Синдром | Без серд | 54 | 73 | 19 | положите | 1 | ж | 15 | 49 | 49 | 1,50 |
| 7 | 216 | аритмия | Дизпласт | Без серд | 65 | 73 | 9 | положите | 1 | ж | 15 | 76 | 76 | 1,50 |
| 8 | 220 | Миокарди | Миокарди | Без серд | 67 | 80 | 13 | отрицател | 1 | ж | 15 | 52 | 52 | 1,75 |
| 9 | 238 | аритмия | Дизпласт | Без серд | 57 | 66 | 9 | положите | 1 | м | 17 | 111 | 111 | 1,71 |
| 10 | 287 | Миокарди | Миокарди | Без серд | 59 | 74 | 16 | положите | 1 | м | 13 | 66 | 66 | 1,85 |
| 11 | 309 | стеноз ао | Пороки с | Без серд | 57 | 60 | 3 | положите | 1 | м | 18 | 65 | 65 | 1,69 |
| 12 | 315 | Коарктаци | Пороки с | Без серд | 31 | 47 | 16 | положите | 1 | ж | 17 | 79 | 79 | 1,63 |
| 13 | 339 | стеноз ле | Пороки с | Без серд | 66 | 67 | 1 | положите | 2 | м | 15 | 63 | 63 | 1,55 |
| 14 | 340 | стеноз ле | Пороки с | Без серд | 68 | 77 | 10 | положите | 1 | м | 16 | 59 | 59 | 1,61 |
| 15 | 348 | стеноз ао | Пороки с | Без серд | 69 | 78 | 9 | положите | 1 | м | 17 | 62 | 62 | 1,79 |
| 16 | 358 | стеноз ле | Пороки с | Без серд | 65 | 71 | 6 | положите | 1 | м | 16 | 55 | 55 | 1,46 |

Рисунок 2.4 – Внесенные данные в редактор данных Statistica

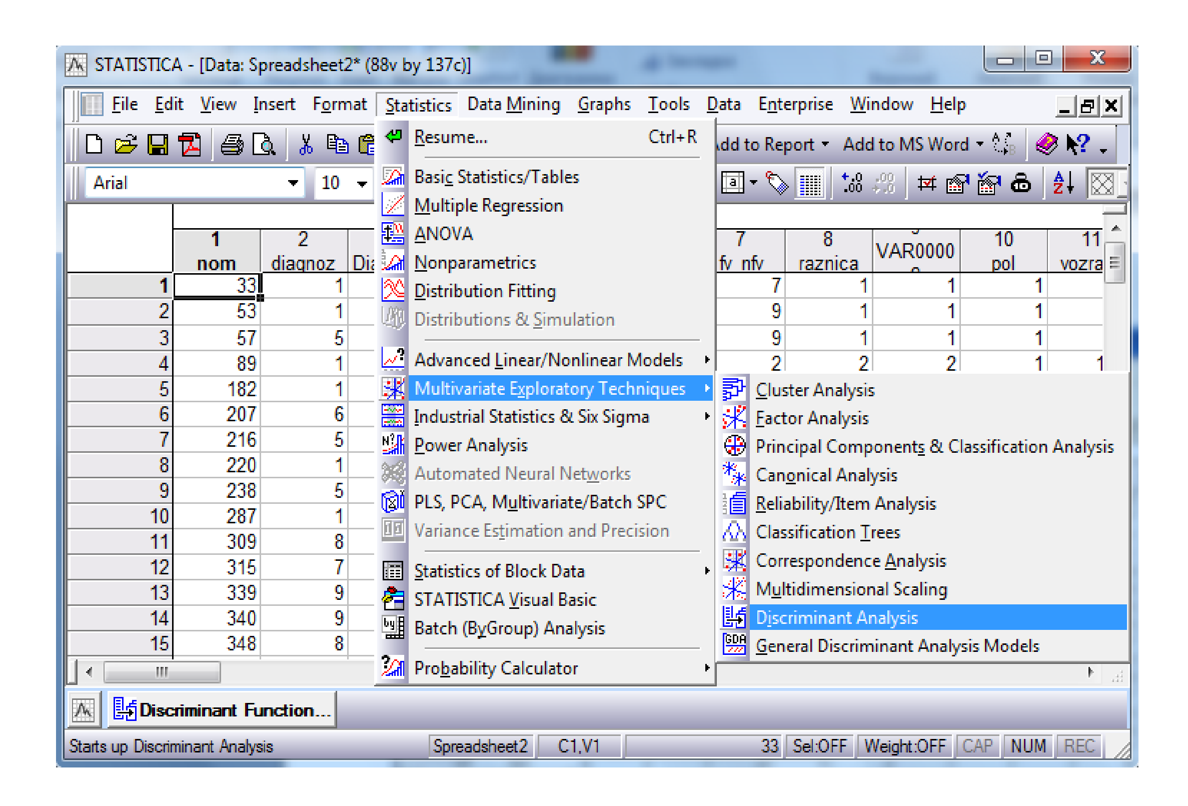



Рисунок 2.5 – Выбор метода анализа данных

В диалоговом окне «Discriminant Function Analysis» ( рис. 2.6) нажали на кнопку «Variables». В открывшемся окне «Select one grouping var. and independent variable list» выбрали группирующую переменную «raznica» и необходимые независимые переменные (рис. 2.7) и нажали OK.

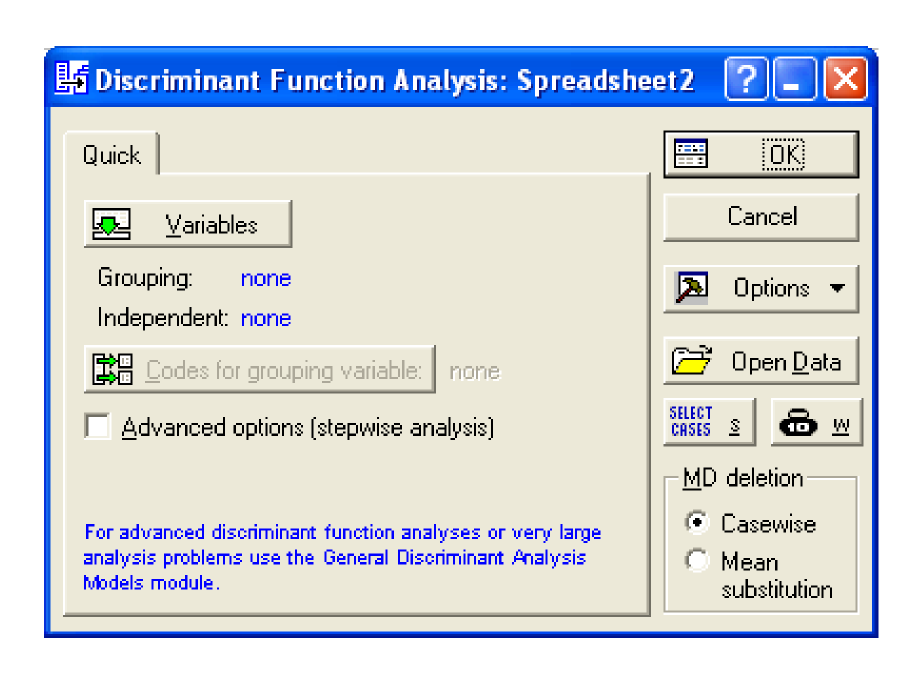



Рисунок 2.6 – Диалоговое окно «Discriminant Function Analysis»

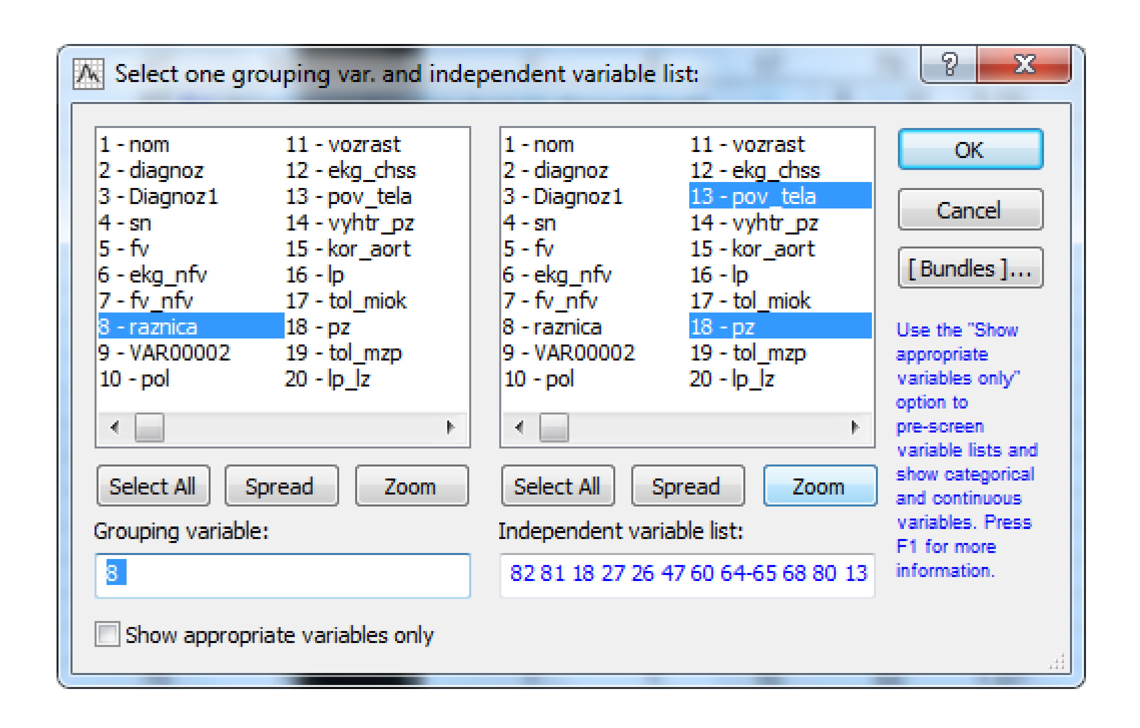



Рисунок 2.7 – Диалоговое окно «Select one grouping var. and independent variable list»

В открывшемся окне (рис. 2.8) «Variables contain text values/text labels» выбрали из предложенного перечня пункт «Continue with current selection».

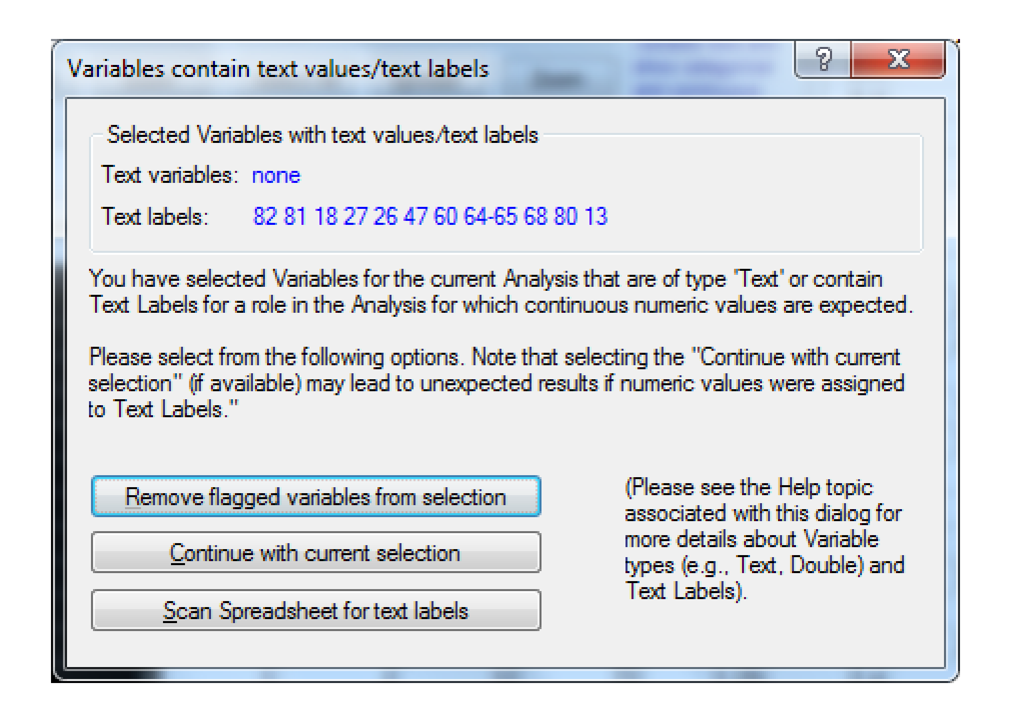



Рисунок 2.8 – Диалоговое окно «Variables contain text values/text labels»

В диалоговом окне «Discriminant Function Analysis» (рис. 2.6) нажали на кнопку «Codes for grouping variable». В появившемся окне «Select codes for grouping variable» выбрали все 3 сравниваемые группы с помощью нажатия кнопки **All** (рис. 2.9) и нажали OK.

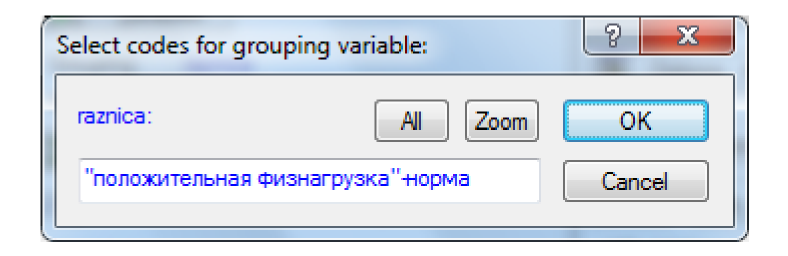



Рисунок 2.9 – Диалоговое окно «Select codes for grouping variable»

В диалоговом окне «Discriminant Function Analysis» (рис. 2.6) установили «галочку» на опции *Advance options (stepwise analysis)* и нажали на кнопку OK, открылось окно диалога «Model Definition» (рис. 2.10).

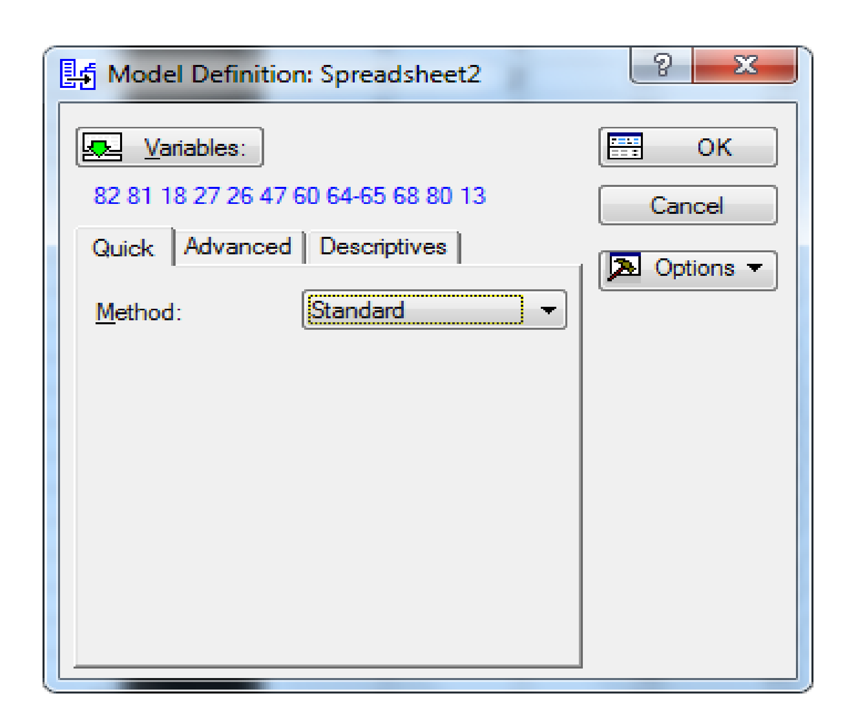



Рисунок 2.10 – Диалоговое окно «Model Definition»

Перейдя на вкладку «**Advanced**» (рис. 2.11), в поле «**Method**» указали метод дискриминантного анализа:- *Forward stepwise* (пошаговый вперед).

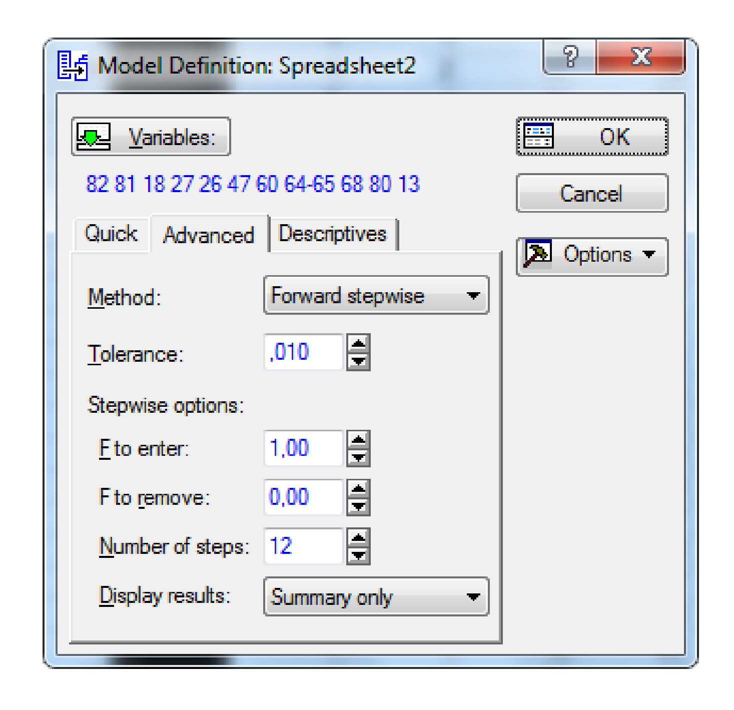



Рисунок 2.11 – Диалоговое окно «Model Definition». Вкладка «Advanced»

При данном выборе программа на последовательных шагах включает переменные в модель.

Опция *Tolerance* (толерантность) задает нижнюю границу толерантности. Толерантность является мерой избыточности переменные. Чем меньше значение толерантности, тем избыточнее переменная в модели, так как переменная несет малую дополнительную информацию [31]. Переменные с толерантностью меньше заданного значения в модель не включаются.

Тест толерантности может обеспечить точность вычислений. Толерантность еще не отобранной переменной равна единице минус квадрат множественной корреляции $R^2$ между этой переменной и всеми уже отобранными переменными, когда корреляции определяются по внутригрупповой корреляционной матрице.

Если проверяемая переменная является линейной комбинацией (или приблизительно равна линейной комбинации) одной или нескольких отобранных переменных, то ее толерантность равна нулю (или близка к нулю). Переменная с малой толерантностью (скажем, меньше 0,001) может привести к ошибке при вычислении матрицы, обратной $E$, ввиду быстрого накопления ошибок округления. Помимо вычислительных проблем, нежелательно использовать переменную, которая является линейной

комбинацией отобранных переменных, потому что она не дает никакой новой информации.

Статистика *F*-включения представляет собой частную *F*-статистику, оценивающую улучшение различения от использования рассматриваемой переменной по сравнению с различением, достигнутым с помощью других уже отобранных переменных. Если величина статистики *F*-включения мала, то такая переменная не будет отобрана, потому что она не дает достаточно большого вклада в различение. Переменная должна пройти проверку толерантности и *F*-включения, прежде чем она будет рассмотрена в соответствии с критерием отбора.

Статистика *F*-удаления также является частной *F*-статистикой. Однако она оценивает значимость ухудшения различения после удаления переменной из списка уже отобранных переменных. Эта процедура проводится в начале каждого шага, чтобы проверить, имеется ли какая-нибудь переменная, уже не вносящая достаточно большого вклада в различение, поскольку отобранные позже переменные дублируют ее вклад. На заключительном шаге статистика *F*-удаления может быть использована для ранжирования дискриминантных возможностей отобранных переменных. Переменная с наибольшим значением статистики *F*-удаления даст наибольший вклад в различение, достигнутое благодаря другим переменным. Переменная, имеющая вторую по величине статистику *F*-удаления, является второй по значимости и т. д.

После выбора в диалоговом окне «Model Definition» метода *Forward stepwise* и нажатия по ОК, открылось окно результатов (рис. 2.12).

Информационная часть окна «Discriminant Function Analysis Results» сообщает, что:

- *Number of variables in model* (число переменных в модели) равно 12;

- *Wilks Lambda* (значение лямбда Уилкса) равно 0,1070642;

- *approx.F* Приближенное значение *F*-статистики с числом степеней свободы 2 и 123 равно 1,725747, а с числом степеней свободы 24 и 246 – 21,07576.

Статистика лямбда Уилкса (λ) вычисляется как отношение детерминанта (определителя) матрицы внутригрупповых ковариаций к детерминанту общей ковариационной матрицы. Значения λ принадлежат интервалу [0,1]. Значения λ, лежащие около 0, свидетельствуют о хорошей дискриминации [32 – 33].

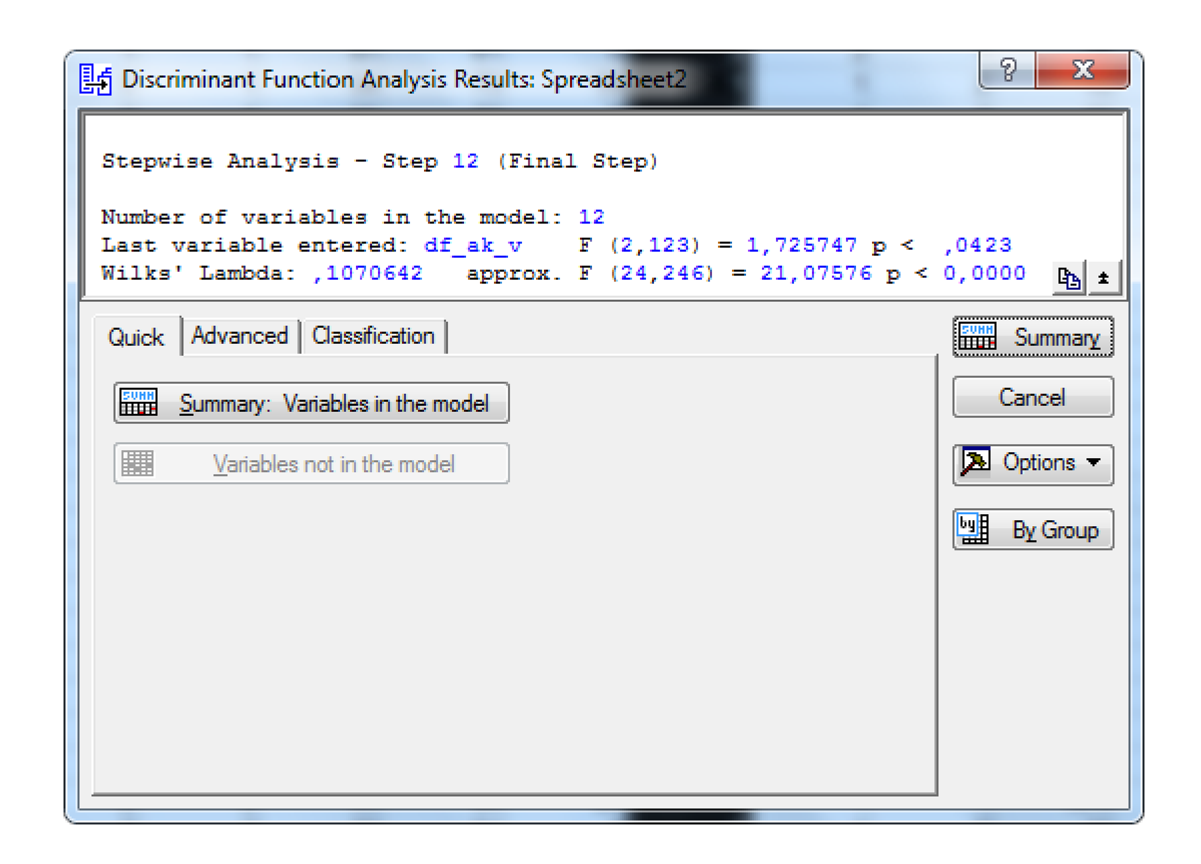



Рисунок 2.12 – Диалоговое окно «Discriminant Function Analysis Results»

Далее в диалоговом окне «Discriminant Function Analysis Results» (рис. 2.12) перешли на вкладку «Advanced» и нажали на кнопку «Summary: Variables in the model» (итоги: переменные, включенные в модель) (рис. 2.13). Появится итоговая таблица анализа данных (рис. 2.14).

После нажатия на кнопку «**Summary: Variables in the model**» (итоги: переменные, включенные в модель) появляется итоговая таблица анализа данных (рис. 2.13).

В первом столбце таблицы приведены значения *Wilks Lambda (лямбда Уилкса),* являющиеся результатом исключения соответствующей переменной из модели. Чем больше значение λ, тем более желательно присутствие этой переменной в процедуре дискриминации.

Значение *Partial Lambda* (частная лямбда) есть отношение лямбда Уилкса после добавления соответствующей переменной к лямбде Уилкса до добавления этой переменной. Частная лямбда характеризует единичный вклад соответствующей переменной в разделительную силу модели. Чем

меньше статистика лямбда Уилкса, тем больше вклад в общую дискриминацию.

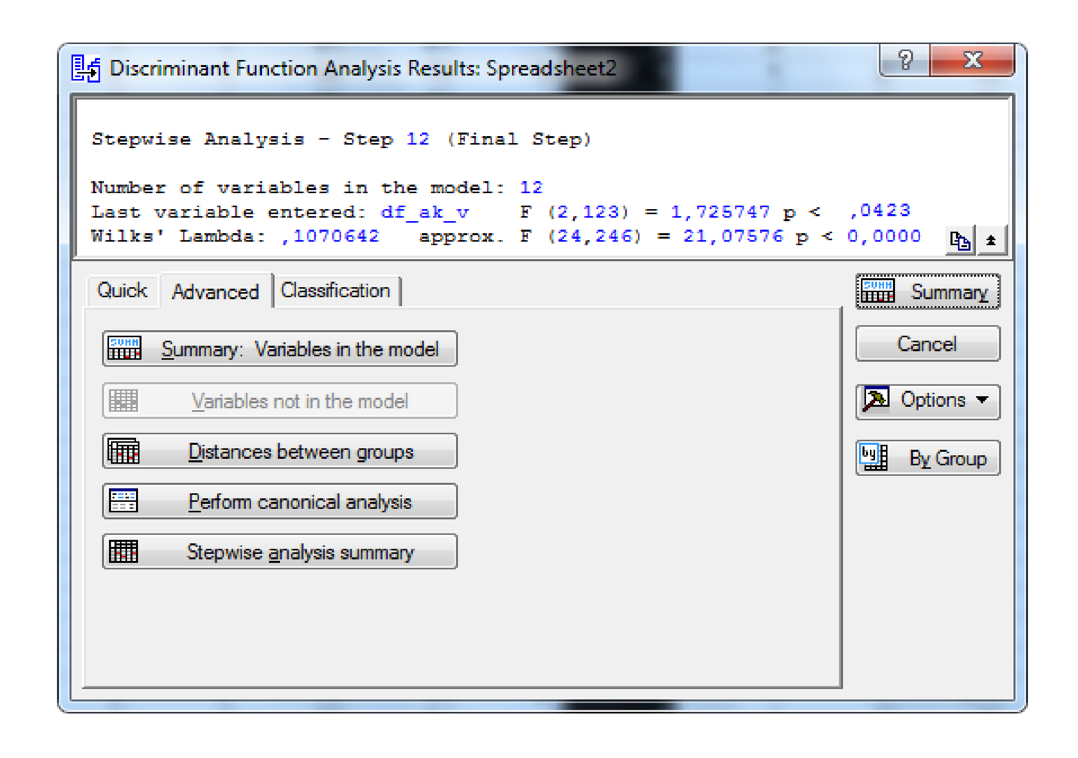



Рисунок 2.13 – Диалоговое окно «Discriminant Function Analysis Results». Вкладка «Advanced»

| N=137 | Discriminant Function Analysis Summary (Spreadsheet2)<br>Step 12, N of vars in model: 12; Grouping: raznica (3 grps)<br>Wilks' Lambda: ,10706 approx. F (24,246)=21,076 p<0,0000 | | | | | |
|---|---|---|---|---|---|---|
| | **Wilks' Lambda** | Partial Lambda | F-remove (2,123) | p-value | Toler. | 1-Toler. (R-Sqr.) |
| **sd95** | 0,134744 | 0,794573 | 15,90011 | 0,000001 | 0,795839 | 0,204162 |
| proz_ds | 0,144306 | 0,741926 | 21,39232 | 0,000000 | 0,618799 | 0,381202 |
| pov_tela | 0,153457 | 0,697680 | 26,64925 | 0,000000 | 0,689161 | 0,310839 |
| pz | 0,136804 | 0,782612 | 17,08298 | 0,000000 | 0,702534 | 0,297466 |
| raas_ang | 0,131900 | 0,811705 | 14,26640 | 0,000003 | 0,900183 | 0,099818 |
| im_fnoa | 0,134035 | 0,798776 | 15,49279 | 0,000001 | 0,713887 | 0,286113 |
| prostats | 0,131742 | 0,812684 | 14,17521 | 0,000003 | 0,840082 | 0,159918 |
| ikso | 0,127706 | 0,838368 | 11,85678 | 0,000020 | 0,646519 | 0,353481 |
| tsamf | 0,114789 | 0,932707 | 4,43713 | 0,013782 | 0,746399 | 0,253601 |
| km_norad | 0,113107 | 0,946578 | 3,47088 | 0,034168 | 0,920736 | 0,079264 |
| raas_ren | 0,112980 | 0,947643 | 3,39789 | 0,036613 | 0,891456 | 0,108544 |
| df_ak_v | 0,110069 | 0,972705 | 1,72575 | 0,182323 | 0,854333 | 0,145667 |

Рисунок 2.14 – Итоги: переменные, включенные в модель

*F-remove* (F-удаление) - это значения *F-критерия,* связанные с соответствующей частной лямбдой Уилкса. Значения p-level – это уровни значимости критериев *F-remove.*

Наибольший вклад в различение дисфункции миокарда дает поверхность тела пациента (рис.2.14).

**Этап 5.** Построение диагностирующей модели для дисфункции миокарда.

В диалоговом окне «Discriminant Function Analysis Results» (рис. 2.13) нажали кнопку «Perform canonical analysis» (выполнение канонического анализа), программа вычислила независимые (ортогональные) дискриминантные функции.

В открывшемся диалоговом окне «Canonical Analysis» (рис. 2.15) нажали кнопку «Coefficients for canonical variables» (коэффициенты канонических переменных).

В результате появились две таблицы с коэффициентами дискриминантных (канонических) функций. В первой таблице представлены исходные (нестандартизированные) коэффициенты дискриминантных функций (рис. 2.16, а). Эти коэффициенты могут быть использованы для вычисления значений канонических переменных для каждого наблюдения каждой дискриминантной функции.

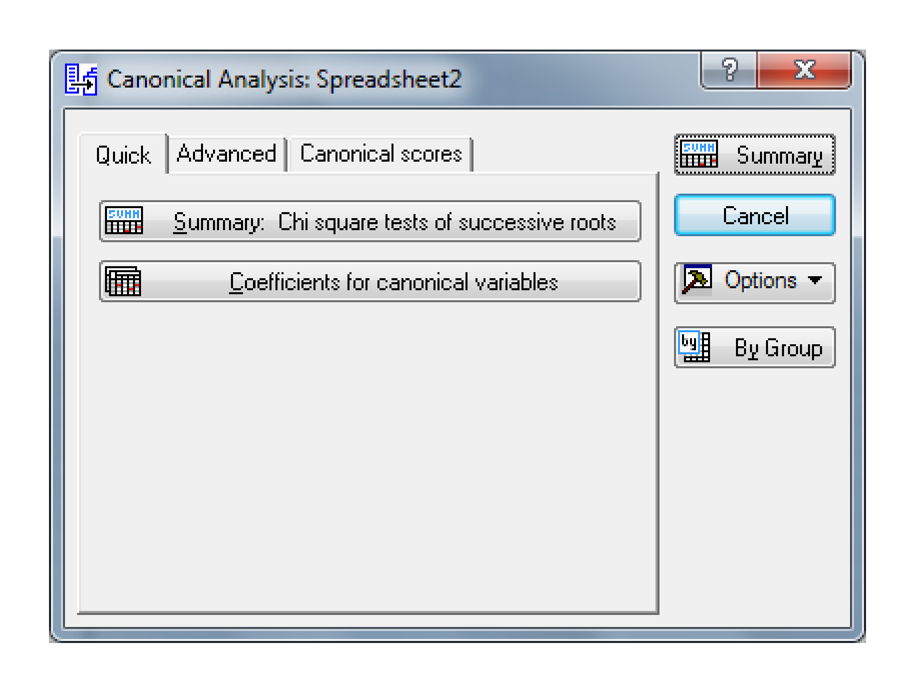



Рисунок 2.15 – Диалоговое окно «Canonical Analysis»

Во второй таблице (рис. 2.16, б) приведены стандартизованные коэффициенты дискриминантных функций. Эти коэффициенты,

основанные на стандартизованных переменных, принадлежат к одной и той же шкале измерений (абсолютной), поэтому их можно сравнивать, чтобы определить величины и направления вкладов переменных в каждую каноническую функцию.

| Variable | Raw Coefficients (Spreadsheet2) for Canonical Variables | |
|---|---|---|
| | **Root 1** | Root 2 |
| **sd95** | -0,11884 | 0,1521 |
| proz_ds | 0,09130 | 0,1368 |
| pov_tela | 2,36799 | 3,7942 |
| pz | -1,64436 | 0,2814 |
| raas_ang | -0,02250 | 0,0008 |
| im_fnoa | -0,06155 | -0,0227 |
| prostats | 0,03467 | -0,0044 |
| ikso | -0,05744 | 0,0525 |
| tsamf | 0,02975 | -0,0131 |
| km_norad | 0,00470 | 0,0069 |
| raas_ren | -0,01913 | 0,7181 |
| df_ak_v | -0,00406 | 0,0031 |
| Constant | -1,21426 | -16,3788 |
| Eigenval | 3,97651 | 0,8769 |
| Cum.Prop | 0,81933 | 1,0000 |

а)

| Variable | Standardized Coefficients (Spreadsheet2) for Canonical Variables | |
|---|---|---|
| | **Root 1** | Root 2 |
| **sd95** | -0,406257 | 0,519831 |
| proz_ds | 0,474964 | 0,711933 |
| pov_tela | 0,468517 | 0,750688 |
| pz | -0,617032 | 0,105605 |
| raas_ang | -0,511460 | 0,017765 |
| im_fnoa | -0,571693 | -0,210560 |
| prostats | 0,525732 | -0,067353 |
| ikso | -0,458566 | 0,418884 |
| tsamf | 0,318365 | -0,140084 |
| km_norad | 0,179144 | 0,263253 |
| raas_ren | -0,009438 | 0,354346 |
| df_ak_v | -0,172402 | 0,132475 |
| Eigenval | 3,976512 | 0,876854 |
| Cum.Prop | 0,819331 | 1,000000 |

б)

Рисунок 2.16 – Коэффициенты дискриминантных функций:
а) нестандартизированные и б) стандартизированные

Стандартизованные коэффициенты показывают вклад переменных в значение дискриминантной функции. Это является одним из подходов к определению значимости переменной. Однако этот подход имеет серьезное ограничение. Если две переменных сильно коррелированны, то их вклад в дискриминантное значение должен разделяться, даже при значительном совместном вкладе. Соответственно их стандартизованные коэффициенты могут быть меньше по сравнению с теми случаями, когда используется одна из этих переменных. Или, другими словами, вклад одного коэффициента частично погашается отрицательным вкладом другого. Это происходит из-за того, что в стандартизованных коэффициентах одновременно принимается во внимание влияние всех переменных.

Таким образом, получены следующие дискриминантные функции:

**DF1** = -0,061* x1 - 0,118* x2+ 0,029* x3 - 0,057* x4 + 0,091* x5 - 0,004*x6 + +2,367* x7 - 0,019* x8 - 0,022* x9+ 0,034* x10-1,644* x11 + 0,005* x12 – –1,214;

**DF2** = – 0,022* x1 + 0,152* x2 – 0,013* x3 + 0,052* x4 + 0,138* x5 + 0,003* * x6 + 3,794 * x7 + 0,718* x8 + 0,0008* x9 – 0,004* x10 + 0,281 * x11 + + 0,007* x12 – 16,378,

где x1 - фактор некроза опухолей (ФНО-а, пг / мл);
x2 - CD-95 (% клеток, несущих маркер апоптоза);
x3 - циклический аденозинмонофосфат (цАМФ, нмоль / мл);
x4 - ИКСО (мл/м2)
x5 - показатель сокращения левого желудочка (%);
x6 - скорость транс-аортального кровотока (см / с);
x7 - поверхность тела (м2);
x8 - ренин (нг / мл / ч);
x9 - ангиотензин-II (пмоль / л);
x10 - простациклин (пг / мл);
x11 - правый желудочек (см);
x12 - норадреналин (нмоль / сут).

В диалоговом окне «Canonical Analysis» выбрали вкладку «Advanced» (рис. 2.17), нажали кнопку «Factor structure» (факторная структура). В появившейся таблице (рис. 2.18) приведены объединенные внутригрупповые корреляции переменных с соответствующими дискриминантными функциями.

Эти корреляции называют еще структурными коэффициентами. Обычно структурные коэффициенты используют для содержательной интерпретации функций, в отличие от коэффициентов дискриминантной функции, которые обозначают вклад каждой переменной в функции. Структурный коэффициент можно вычислить и для переменной в пределах отдельно взятой группы. Тогда получаем внутригрупповой структурный коэффициент $z_{il}$ для *i*-й переменной *l*-й функции, который вычисляется по формуле:

$$z_{il} = \sum_{j=1}^{p} r_{ij} c_{jl} = \sum_{j=1}^{p} \frac{e_{ij} c_{jl}}{\sqrt{e_{ii} e_{ll}}},$$

где $r_{ij}$ - внутригрупповые структурные коэффициенты корреляции между переменными $i$ и $j$; $c_{jl}$ - стандартизованные коэффициенты канонической функции для переменной $j$ и функции $l$ [20].

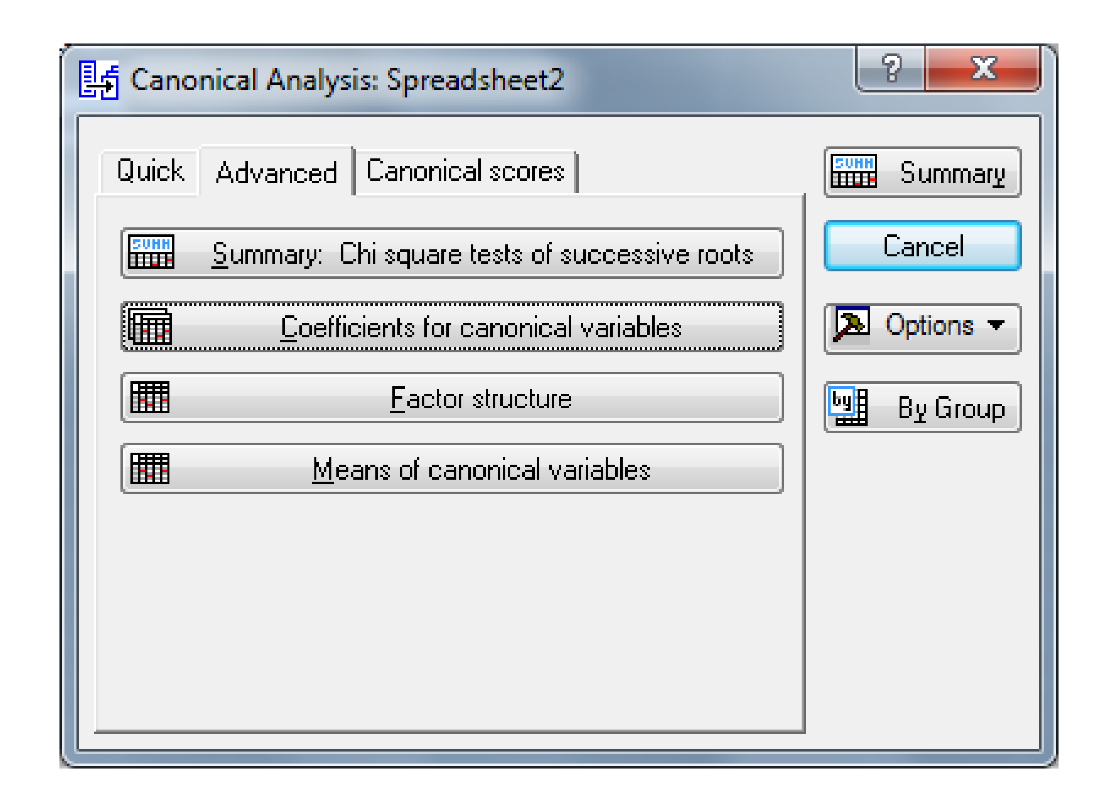



Рисунок 2.17 – Диалоговое окно «Canonical Analysis». Вкладка «Advanced»

С другой стороны, структурные коэффициенты являются просто двуместными корреляциями, поэтому на них не влияют взаимные зависимости прочих переменных.

В работе [34] описывается, как структурные коэффициенты могут использоваться для графического представления различия между групповыми центроидами в случае двух канонических дискриминантных функций. На графике с осями, которые относятся к этим двум функциям, представлены групповые центроиды и главный центроид, изображены векторы, исходящие из главного центроида и направленные в каждую дискриминантную переменную. Направляющие: углы этих векторов

вычисляются, исходя из структурных коэффициентов. Длина вектора определяется межгрупповыми и внутригрупповыми вариациями соответствующей переменной.

| Variable | Factor Structure Matrix (Spreadsheet2) Correlations Variables - Canonical Roots (Pooled-within-groups correlations) Root 1 | Root 2 |
|---|---|---|
| sd95 | -0,351153 | 0,176843 |
| proz_ds | 0,318809 | 0,308907 |
| pov_tela | 0,057711 | 0,532739 |
| pz | -0,169708 | 0,348437 |
| raas_ang | -0,300870 | 0,088785 |
| im_fnoa | -0,284637 | -0,061710 |
| prostats | 0,243468 | -0,091718 |
| ikso | -0,242858 | 0,080039 |
| tsamf | 0,003884 | -0,088459 |
| km_norad | 0,012844 | 0,253040 |
| raas_ren | 0,006142 | 0,270468 |
| df_ak_v | -0,085334 | 0,168433 |

Рисунок 2.18 – Объединенные внутригрупповые корреляции переменных с соответствующими дискриминантными функциями

Полученная диаграмма дает наглядное представление о различиях групп с помощью дискриминантных переменных, а также о потенциальных возможностях этих переменных.

В окне «Canonical Analysis» нажали на кнопку «Summary: Chi square tests of successive roots» (итоги: хи-квадрат критерий последовательности корней) (рис. 2.15). Появится таблица результатов с пошаговым критерием для канонических корней — дискриминантных функций (рис. 2.19).

| Roots Removed | Chi-Square Tests with Successive Roots Removed (Spre Eigen-value | Canonicl R | Wilks' Lambda | Chi-Sqr. | df | p-value |
|---|---|---|---|---|---|---|
| 0 | 3,976512 | 0,893899 | 0,107064 | 287,1109 | 24 | 0,000000 |
| 1 | 0,876854 | 0,683516 | 0,532806 | 80,9032 | 11 | 0,000000 |

Рисунок 2.19 – Таблица результатов с пошаговым критерием для канонических корней

В диалоговом окне «Canonical Analysis» перешли на вкладку «Canonical scores» (канонические значения) и нажали кнопку «Canonical scores for each case» (канонические значения для каждого наблюдения)

(рис. 2.20). Появится таблица (рис. 2.21) со значениями дискриминантных функций для каждого наблюдения.

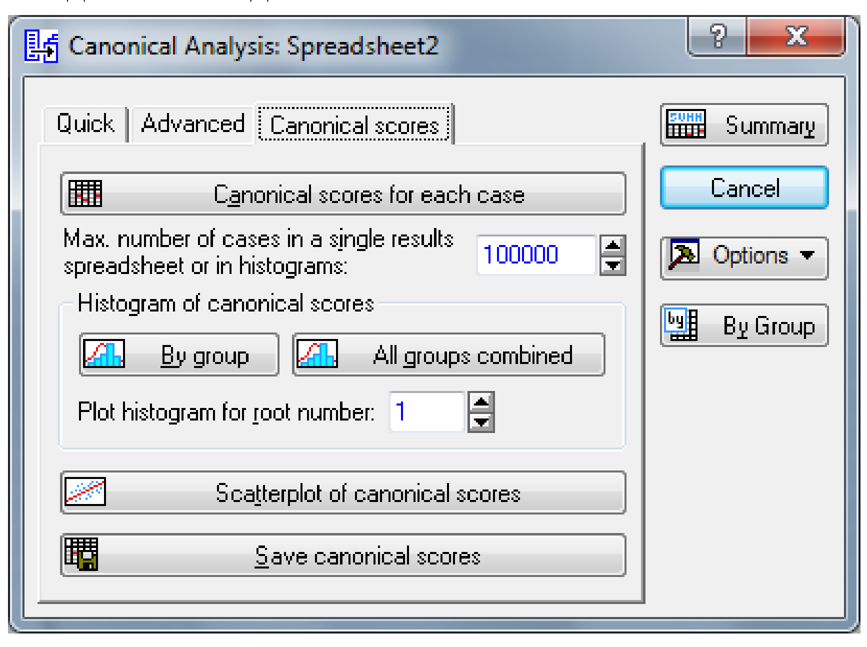



Рисунок 2.20 – Диалоговое окно «Canonical Analysis». Вкладка «Canonical scores»

Unstandardized Canonical Scores (Spreadsheet2)

| Case | Group | Root 1 | Root 2 |
|---|---|---|---|
| 1 | положительная физнагрузка | -1,77625 | 0,30163 |
| 2 | положительная физнагрузка | -4,57447 | -0,79041 |
| 3 | положительная физнагрузка | -4,31024 | -1,24178 |
| 4 | отрицательная физнагрузка | -0,90177 | 2,46267 |
| 5 | отрицательная физнагрузка | -1,67667 | 1,17803 |
| 6 | положительная физнагрузка | -0,94063 | -1,28100 |
| 7 | положительная физнагрузка | -3,71312 | -0,28700 |
| 8 | отрицательная физнагрузка | 0,16817 | 0,57280 |
| 9 | положительная физнагрузка | -3,43594 | 1,19067 |
| 10 | положительная физнагрузка | -0,59223 | 0,52584 |
| 11 | положительная физнагрузка | -2,83045 | -0,73086 |
| 12 | положительная физнагрузка | -2,52023 | -0,68473 |
| 13 | положительная физнагрузка | -0,52004 | 0,04671 |
| 14 | положительная физнагрузка | -2,72984 | 0,12341 |
| 15 | положительная физнагрузка | -2,97920 | -1,61743 |
| 16 | положительная физнагрузка | -1,63367 | -0,43257 |
| 17 | положительная физнагрузка | -4,44589 | -0,78120 |
| ... | ... | ... | ... |
| 136 | норма | 2,53059 | -0,64247 |
| 137 | норма | 1,67824 | 0,27540 |

Рисунок 2.21 – Таблица со значениями дискриминантных функций для каждого наблюдения

В окне «Canonical Analysis» нажали кнопку «Scatterplot of canonical scores» (диаграмма рассеяния для канонических значений). Появилась

диаграмма рассеяния (рис. 2.22) канонических значений для пар значений дискриминантных функций. На диаграмме видно, что наблюдения, принадлежащие одинаковым группам, локализованы в определенных областях плоскости.

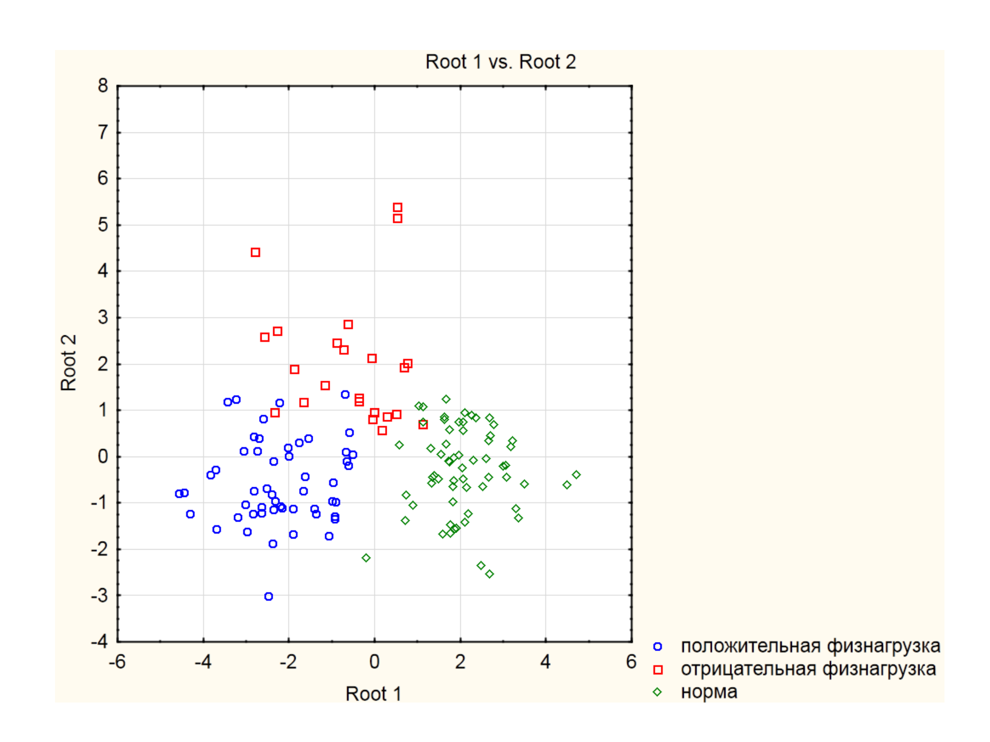



Рисунок 2.22 – Диаграмма рассеяния исследуемых групп пациентов

Вернувшись в окно результатов «Discriminant Function Analysis Results», нажали кнопку «Cancel», и активизировали вкладку «Classification» (классификация). Открылось окно результатов классификации (рис. 2.23).

В рамке *A priori classification probabilities* (априорные вероятности классификации) приведены различные опции задания априорных вероятностей того, что наблюдение при классификации попадет в одну из групп: *Proportional to group sizes* (пропорциональные размерам групп); *Same for alt groups* (одинаковые для всех групп); *User defined* (заданные пользователем). Априорные вероятности могут существенно влиять на точность классификации. Если есть предварительные сведения (оценки) о возможном количественном соотношении наблюдений в группах, то желательно выбрать опцию *User defined*, если таких сведений нет и число наблюдений в группах примерно одинаково, то надо выбрать *Same for all groups*, в противном случае - *Proportional to group sees*.

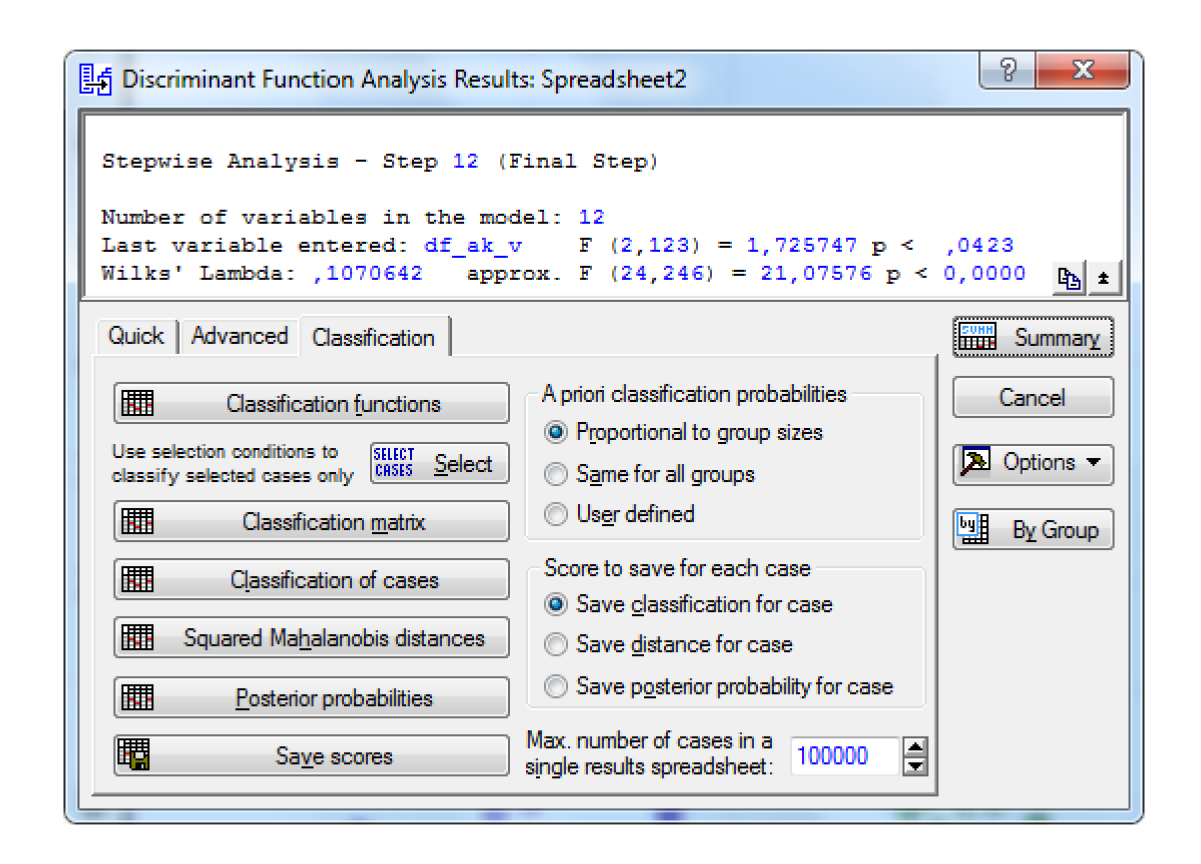



Рисунок 2.23 – Диалоговое окно «Discriminant Function Analysis Results». Вкладка «Classification»

В рамке *Score to save for each case* (сохранить для каждого наблюдения) приведены опции, при выборе которых можно сохранить тот или иной результат классификации: *Save classification for case* (сохранить результаты классификации); *Save distance for case* (сохранить расстояния); *Save posterior probability for case* (сохранить апостериорные вероятности).

В строке *Max. number of cases in a single results spreadsheet* (максимальное число наблюдений в таблице результатов). Если наблюдений больше указанного числа, то результаты будут выделены несколькими таблицами.

В диалоговом окне «Discriminant Function Analysis Results» на вкладке «Classification» нажали кнопку «Classification functions» (функции классификации). Функции классификации - это линейные функции, которые вычисляются для каждой группы и могут быть использованы для классификации наблюдений (пациентов). Исследуемого пациента приписывают к той группе, для которой классификационная функция имеет наибольшее значение.

В таблице, изображенной на рис. 2.24, приведены коэффициенты и свободные члены при переменных линейных функций. Например, классификационные уравнения для всех 3-х групп имеют вид:

**Группа1** = - x1* 0,292+ x2 * 3,588+ x3 * 0,089+ x4 * 1,202+ x5* *2,546+ x6* 0,060+ x7* 61,200+ x8* 6,133+ x9* 0,067+ x10* 0,253 + x11* *5,770 + x12* 0,049 - 159,393.

**Группа2** = - x1* 0,452+ x2 * 3,789+ x3 * 0,104+ x4 * 1,244+ x5* *3,052+ x6* 0,061+ x7* 74,958+ x8* 7,971+ x9* 0,032+ x10* 0,298 + x11* *3,807 + x12* 0,075 - 204,465.

**Группа3** = - x1* 0,563+ x2 * 3,118+ x3 * 0,213+ x4 * 0,969 + x5* *2,975+ x6* 0,043+ x7* 72,414+ x8* 6,245- x9* 0,029+ x10* 0,401 - x11* *1,228 + x12* 0,071 - 168,451

где группа1 - пациенты с адекватной реакцией на физическую нагрузку; группа2 - пациенты с неадекватной реакцией на физическую нагрузку; группа3 – контрольная группа .

| | Classification Functions; grouping: raznica (S | | |
|---|---|---|---|
| Variable | **положительная физнагрузка p=,37956** | отрицательная физнагрузка p=,16788 | норма p=,45255 |
| **sd95** | 3,588 | 3,789 | 3,118 |
| proz_ds | 2,546 | 3,052 | 2,975 |
| pov_tela | 61,200 | 74,958 | 72,414 |
| pz | 5,770 | 3,807 | -1,228 |
| raas_ang | 0,067 | 0,032 | -0,029 |
| im_fnoa | -0,292 | -0,452 | -0,563 |
| prostats | 0,253 | 0,298 | 0,401 |
| ikso | 1,202 | 1,244 | 0,969 |
| tsamf | 0,089 | 0,104 | 0,213 |
| km_norad | 0,049 | 0,075 | 0,071 |
| raas_ren | 6,133 | 7,971 | 6,245 |
| df_ak_v | 0,060 | 0,061 | 0,043 |
| Constant | -159,393 | -204,465 | -168,451 |

Рисунок 2.24 – Классификационные функции

В случае двух групп разделяющая точка С (константа дискриминации) между двумя группами равна полусумме величин двух центроидов этих групп.

В случае трех групп разделяющие точки С1, С2 , …, С 3.

Правило классификации имеет вид: если f > С 1 – пациент принадлежит первой группе; С1 < f < С2 – второй группе; f < С 3 – 3-й группе.

В диалоговом окне «Discriminant Function Analysis Results» (рис. 2.23) нажали кнопку «Classification of cases» (классификация наблюдений). Программа вывела таблицу классификации для каждого наблюдения (рис. 2.25). Классификации упорядочены по первому, второму и третьему выбору. Столбец 1 содержит первый классификационный выбор, т.е. группу, для которой соответствующее наблюдение имеет наивысшую апостериорную вероятность и наибольшее значение классификационной функции. Наблюдения, которые не удалось правильно классифицировать, отмечены *.

| Case | Classification of Cases (Spreadsheet2)<br>Incorrect classifications are marked with *<br>Observed Classif. | 1<br>p=,37956 | 2<br>p=,16788 | 3<br>p=,45255 |
|---|---|---|---|---|
| 1 | положительная физнагрузка | положительная физнагрузка | отрицательная физнагрузка | норма |
| 2 | положительная физнагрузка | положительная физнагрузка | отрицательная физнагрузка | норма |
| 3 | положительная физнагрузка | положительная физнагрузка | отрицательная физнагрузка | норма |
| 4 | отрицательная физнагрузка | отрицательная физнагрузка | положительная физнагрузка | норма |
| * 5 | отрицательная физнагрузка | положительная физнагрузка | отрицательная физнагрузка | норма |
| 6 | положительная физнагрузка | положительная физнагрузка | норма | отрицательная физнагрузка |
| 7 | положительная физнагрузка | положительная физнагрузка | отрицательная физнагрузка | норма |
| * 8 | отрицательная физнагрузка | норма | отрицательная физнагрузка | положительная физнагрузка |
| 9 | положительная физнагрузка | положительная физнагрузка | отрицательная физнагрузка | норма |
| 10 | положительная физнагрузка | положительная физнагрузка | отрицательная физнагрузка | норма |
| 11 | положительная физнагрузка | положительная физнагрузка | отрицательная физнагрузка | норма |
| 12 | положительная физнагрузка | положительная физнагрузка | отрицательная физнагрузка | норма |
| 13 | положительная физнагрузка | положительная физнагрузка | отрицательная физнагрузка | норма |
| 14 | положительная физнагрузка | положительная физнагрузка | отрицательная физнагрузка | норма |
| … | … | … | … | … |
| 132 | норма | норма | положительная физнагрузка | отрицательная физнагрузка |
| 133 | норма | норма | отрицательная физнагрузка | положительная физнагрузка |
| 134 | норма | норма | отрицательная физнагрузка | положительная физнагрузка |
| 135 | норма | норма | отрицательная физнагрузка | положительная физнагрузка |
| 136 | норма | норма | отрицательная физнагрузка | положительная физнагрузка |
| 137 | норма | норма | отрицательная физнагрузка | положительная физнагрузка |

Рисунок 2.25 – Таблица классификации для каждого наблюдения

На вкладке «Classification» в диалоговом окне «Discriminant Function Analysis Results» (рис. 2.23) нажали кнопку «Squared Mahalanobis distances»

(квадрат расстояния Махаланобиса). Была выведена таблица квадратов расстояний Махалонобиса каждого наблюдения от центроида группы (рис. 2.26).

| Case | Squared Mahalanobis Distances from Group Centroids (Spreadsheet2) Incorrect classifications are marked with * Observed Classif. | положительная физнагрузка p=,37956 | отрицательная физнагрузка p=,16788 | норма p=,45255 |
|---|---|---|---|---|
| 1 | положительная физнагрузка | 8,3923 | 11,91088 | 22,67144 |
| 2 | положительная физнагрузка | 17,3010 | 35,68066 | 56,24844 |
| 3 | положительная физнагрузка | 14,5754 | 34,43840 | 51,49346 |
| 4 | отрицательная физнагрузка | 40,7414 | 30,14211 | 46,32682 |
| * 5 | отрицательная физнагрузка | 12,8190 | 11,44853 | 25,76708 |
| 6 | положительная физнагрузка | 15,1661 | 24,18356 | 23,11127 |
| 7 | положительная физнагрузка | 11,2858 | 24,22021 | 42,54937 |
| * 8 | отрицательная физнагрузка | 10,5497 | 6,28034 | 7,95111 |
| 9 | положительная физнагрузка | 12,5423 | 16,87510 | 40,62140 |
| 10 | положительная физнагрузка | 9,5842 | 8,05292 | 13,55404 |
| 11 | положительная физнагрузка | 7,6272 | 19,97784 | 31,53586 |
| 12 | положительная физнагрузка | 40,4120 | 51,50520 | 61,62639 |
| 13 | положительная физнагрузка | 6,7096 | 7,43588 | 10,31746 |
| 14 | положительная физнагрузка | 6,8642 | 14,43755 | 29,44491 |
| ... | ... | ... | ... | ... |
| 132 | норма | 29,5131 | 32,87676 | 8,64295 |
| 133 | норма | 34,2684 | 25,79112 | 7,22740 |
| 134 | норма | 16,1526 | 14,02534 | 4,20093 |
| 135 | норма | 20,5923 | 16,28628 | 1,60706 |
| 136 | норма | 29,0732 | 23,38361 | 6,80447 |
| 137 | норма | 19,3685 | 11,69558 | 3,93736 |

Рисунок 2.26 – Таблица квадратов расстояний Махалонобиса каждого наблюдения от центроида группы

Эти расстояния аналогичны квадратам евклидовых расстояний, но учитывают корреляции между переменными в модели. Пациента приписывают группе, к которой он ближе всего. Пациенты, которых не удалось правильно классифицировать, также отмечены *.

**Этап 6.** Определение τ-статистики ошибок и доли правильной диагностики.

На вкладке «Classification» в диалоговом окне «Discriminant Function Analysis Results» (рис. 2.23) нажали кнопку «Classification matrix» (матрица классификации). Матрица (рис. 2.27) содержит информацию о количестве и проценте корректно классифицированных наблюдений (пациентов) в каждой группе. Строки матрицы — исходные классы, столбцы — предсказанные классы.

| Group | Classification Matrix (Spreadsheet2)<br>Rows: Observed classifications<br>Columns: Predicted classifications<br>Percent Correct | положительная физнагрузка p=,37956 | отрицательная физнагрузка p=,16788 | норма p=,45255 |
|---|---|---|---|---|
| **положительная физнагрузка** | 98,07692 | 51 | 1 | 0 |
| отрицательная физнагрузка | 78,26087 | 2 | 18 | 3 |
| норма | 98,38710 | 1 | 0 | 61 |
| Total | 94,89051 | 54 | 19 | 64 |

Рисунок 2.27 – Результаты классификации

К группе «Положительная физнагрузка» относится 52 наблюдения. 51 наблюдение (98,07692 %) было определено корректно, 1 ошибочно отнесено к группе «Отрицательная физнагрузка». В группе «Отрицательная физнагрузка» насчитывается 23 наблюдения. 18 из них были определены корректно, 2 наблюдения ошибочно отнесено к группе «Положительная физнагрузка» и 3 к группе «Норма». В группе «Норма» насчитывается 62 наблюдения. Корректно определены 61, ошибочно – 1 к группе «Положительная физнагрузка».

Анализ полученных результатов (рис.2.27) показал, что прогноз для исследуемых групп сделан верно (корректно классифицированы в общей сложности 94,89 % всех случав заболеваний сердца). Данное значение получено отношением корректно определенных наблюдений к общему их количеству (130/137=94, 89). Как видно из данного расчета, не учитывались априорные вероятности каждой группы пациентов.

В пакете STATISTICA есть возможность задавать различные опции априорных вероятностей. В диалоговом окне «Discriminant Function Analysis Results» (рис. 2.22) на вкладке «Classification» нажали кнопку «Posterior probabilities» (апостериорные вероятности). В открывшейся таблице (рис. 2.28) каждому пациенту поставлена в соответствие вероятность принадлежности к группам. Эта вероятность определяется посредством расстояний Махалонобиса и априорных вероятностей. Чем дальше наблюдение расположено от центра группы, тем менее вероятно, что оно принадлежит этой группе. Пациента приписывают к той группе, для которой имеется наибольшая апостериорная вероятность классификации. Априорные вероятности могут быть заданы

пользователем, могут быть равны для всех групп, могут быть пропорциональны размерам групп.

| Case | Observed Classif. | положительная физнагрузка p=,37956 | отрицательная физнагрузка p=,16788 | норма p=,45255 |
|---|---|---|---|---|
| 1 | положительная физнагрузка | 0,928423 | 0,070699 | 0,000878 |
| 2 | положительная физнагрузка | 0,999955 | 0,000045 | 0,000000 |
| 3 | положительная физнагрузка | 0,999978 | 0,000022 | 0,000000 |
| 4 | отрицательная физнагрузка | 0,011154 | 0,988031 | 0,000815 |
| * 5 | отрицательная физнагрузка | 0,532065 | 0,466956 | 0,000979 |
| 6 | положительная физнагрузка | 0,973411 | 0,004741 | 0,021848 |
| 7 | положительная физнагрузка | 0,999313 | 0,000687 | 0,000000 |
| * 8 | отрицательная физнагрузка | 0,109756 | 0,410417 | 0,479827 |
| 9 | положительная физнагрузка | 0,951760 | 0,048239 | 0,000001 |
| 10 | положительная физнагрузка | 0,472827 | 0,449718 | 0,077455 |
| 11 | положительная физнагрузка | 0,999073 | 0,000919 | 0,000008 |
| 12 | положительная физнагрузка | 0,998248 | 0,001722 | 0,000029 |
| 13 | положительная физнагрузка | 0,664919 | 0,204546 | 0,130535 |
| 14 | положительная физнагрузка | 0,990058 | 0,009928 | 0,000015 |
| ... | ... | ... | ... | ... |
| 132 | норма | 0,000025 | 0,000002 | 0,999973 |
| 133 | норма | 0,000001 | 0,000035 | 0,999964 |
| 134 | норма | 0,002120 | 0,002716 | 0,995165 |
| 135 | норма | 0,000063 | 0,000241 | 0,999696 |
| 136 | норма | 0,000012 | 0,000093 | 0,999895 |
| 137 | норма | 0,000371 | 0,007606 | 0,992023 |

Posterior Probabilities (Spreadsheet2)
Incorrect classifications are marked with *

Рисунок 2.28 – Таблица с апостериорными вероятностями

Например, в последней строке (пациент №137) (рис. 2.28) максимальное значение вероятности соответствует группе 3 (норма), т.е. с вероятностью 0,992023 данный пациент относится к группе №3. Согласно таблице квадратов расстояний Махалонобиса (рис. 2.26) расстояние от данного пациента до центроидов групп минимально для группы 3 (3,93736). Это дополнительное подтверждение того, что исследуемый пациент относится к группе «норма».

В случае, когда флажок как на рис. 2.23 установлен в поле «Same for alt groups» (одинаковые для всех групп), каждая группа пациентов имеет априорную вероятность, равную 0,333. Следовательно, число объектов, которые будут правильно предсказаны при случайной классификации их по группам пропорционально априорным вероятностям, равно:

$$\sum_{k=1}^{g} p_k n_k = (0{,}333 \cdot 52) + (0{,}333 \cdot 23) + (0{,}333 \cdot 62) = 45{,}62.$$

Для 130 правильных предсказаний из 137 возможных τ-статистика, согласно формуле (1.9), составит:

$$\tau = \frac{130 - 45,62}{137 - 45,62} = 0,923.$$

Это означает, что классификация с помощью дискриминантных функций делает на 92,3% ошибок меньше, чем ожидалось при случайной классификации.

В случае если флажок установлен в поле «Proportional to group sizes» (пропорциональные размерам групп) (рис. 2. 2), каждая группа пациентов имеет свою априорную вероятность. Следовательно, число объектов, которые будут правильно предсказаны при случайной классификации их по группам пропорционально априорным вероятностям, равно:

$$\sum_{k=1}^{g} p_k n_k = (0,379 \cdot 52) + (0,168 \cdot 23) + (0,453 \cdot 62) = 51,66.$$

Для 130 правильных предсказаний из 137 возможных τ -статистика согласно формуле (1.9) составит:

$$\tau = \frac{130 - 51,66}{137 - 51,66} = 0,918.$$

В случае если флажок установлен в поле «User defined» (рис. 2.22), то априорные вероятности задаются пользователем.

В качестве показателя качества алгоритма диагностики (прогностической «силы») использовали долю правильной диагностики μ, определяемую по формуле (1.10).

$p_1$=52/137=0,379; $p_2$=23/137=0,168; $p_3$=62/137=0,453.

$$\sum_{k=1}^{g} p_k n_k = 0,379 + 0,168 + 0,453 = 1.$$

$$\kappa = \frac{51}{52} = 0,981;$$

$$\lambda = \frac{18}{23} = 0,783;$$

$$\eta = \frac{61}{62} = 0,984.$$

$$\mu = 0,379 \cdot 0,981 + 0,168 \cdot 0,783 + 0,453 \cdot 0,984 = 0,9489.$$

Для корректного применения дискриминантных функции при решении поставленной задачи использовались две выборки пациентов, по 137 человек каждая: одна для вычисления дискриминантной функции (анализируемая), вторая – проверочная (для проверки результатов расчета на основании первой выборки). Т.е была проведена кросс-проверка (перекрестная проверка).

В результате проведения клинической апробации были получены следующие результаты (рис. 2.29). К группе «Положительная физнагрузка» относится 52 наблюдения. 51 наблюдение (98,07692 %) было определено корректно, 1 ошибочно отнесено к группе «Отрицательная физнагрузка». В группе «Отрицательная физнагрузка» насчитывается 23 наблюдения. 20 из них были определены корректно, 2 наблюдения ошибочно отнесено к группе «Положительная физнагрузка» и 1 к группе «Норма». В группе «Норма» насчитывается 62 наблюдения. Корректно определены 61, ошибочно – 1 к группе «Положительная физнагрузка».

| | Classification Matrix (Spreadsheet2)<br>Rows: Observed classifications<br>Columns: Predicted classifications | | | |
|---|---|---|---|---|
| Group | **Percent Correct** | положительная физнагрузка p=,37956 | отрицательная физнагрузка p=,16788 | норма p=,45255 |
| **положительная физнагрузка** | 98,07692 | 51 | 1 | 0 |
| отрицательная физнагрузка | 86,95652 | 2 | 20 | 1 |
| норма | 98,38710 | 1 | 0 | 61 |
| Total | 96,35036 | 54 | 21 | 62 |

Рисунок 2.29 – Результаты классификации

Анализ полученных результатов показал, что прогноз для исследуемых групп  сделан верно (корректно классифицированы в общей сложности 96, 35 % всех случав заболеваний сердца).

Результаты оценки показателя качества алгоритма диагностики (прогностической «силы») приведены ниже.

$p_1$=52/137=0,379;  $p_2$=23/137=0,168;  $p_3$=62/137=0,453.

$$\sum_{k=1}^{g} p_k n_k = 0,379 + 0,168 + 0,453 = 1.$$

$$\kappa = \frac{51}{52} = 0,981;$$

$$\lambda = \frac{20}{23} = 0,869;$$

$$\eta = \frac{61}{62} = 0,984.$$

$$\mu = 0,379 \cdot 0,981 + 0,168 \cdot 0,870 + 0,453 \cdot 0,984 = 0,9635$$

**Этап 7.** Формирование диагностического заключения.

Таким образом, использование разработанного метода позволяет выявить систолическую дисфункцию миокарда у подростков и своевременными превентивными мерами предотвратить снижение физической активности пациента.

Предметом дальнейшего совершенствования метода является использование в диагностирующей модели других клинических показателей в соответствии с новыми концепциями диагностики систолической дисфункции миокарда.

# ГЛАВА 3. ИСПОЛЬЗОВАНИЕ БИНАРНОЙ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ ДЛЯ ОПРЕДЕЛЕНИЯ СЕРДЕЧНО-СОСУДИСТОГО РИСКА У ПАЦИЕНТОВ С ПОЧЕЧНОЙ НЕДОСТАТОЧНОСТЬЮ

В течение последних десятилетий в мире наблюдается повсеместный неуклонный рост числа больных, страдающих хронической почечной недостаточностью (ХПН). В частности, в Украине насчитывается около 130,8 больных на 100 тыс. населения, а также наблюдается высокая смертность пациентов с данной патологией (8,9%) [35]. Это делает проблему лечения ХПН одной из центральных в современной нефрологии. Значительный прогресс медицины, научно-технические достижения в области гемодиализа, широкое внедрение в клинику перитонеального диализа и возможность трансплантации почки создали реальные предпосылки для успешного решения этой проблемы [36].

Однако, если диагностика и лечение ХПН достаточно хорошо разработаны, то при ведении пациентов с сочетанной патологией часто возникают трудности. Особое место среди сочетанных состояний занимает ишемическая болезнь сердца (ИБС), поскольку пациенты с нарушением функции почек значительно чаще умирают от острой ишемии миокарда, чем от собственно терминальной почечной недостаточности. Не менее 20% всех случаев смерти при ХПН обусловлены инфарктом миокарда (ИМ) [37]. Своевременная профилактика и лечение сердечно-сосудистых заболеваний при ХПН имеет важное значение для улучшения результатов диализа. В этой связи приоритет приобретает исследование факторов риска, влияющих на возникновение, клинические проявления и прогноз сердечно-сосудистых заболеваний у пациентов с патологией почек.

Еще более актуальным представляется решение этих вопросов в свете современной демографической проблемы, а именно в аспекте данных о непрерывном росте в последние 15-20 лет числа больных, страдающих хронической болезнью почек (ХБП), и результатов популяционных исследований, демонстрирующих вклад ХБП в кардиоваскулярную смертность населения.

В настоящее время можно считать доказанным, что снижение функции почек является независимым фактором риска развития сердечно-сосудистой патологии. Почечная недостаточность может быть либо маркером пока неизвестных факторов кардиоваскулярного риска, либо

сама по себе вызывать и ускорять прогрессирование сердечно-сосудистой патологии, следовательно, выявление факторов, влияющих на прогрессирование ИБС у этой категории больных является актуальной задачей.

Для решения задачи определения факторов риска развития сердечно-сосудистой патологии используют дискриминантный анализ, представленный в главах 1 и 2 данной монографии. Применение дискриминантного анализа подтверждается широким распространением пакетов прикладных программ, включающих его.

Несмотря на известные достижения в области классификации и анализа информативности данных средствами дискриминантного анализа имеется ряд сдерживающих факторов. Отдельные измерения могут быть тесно связанными (коррелированными). Все это приводит к вырожденности или плохой обусловленности оценок ковариационных матриц (плохо обусловленные задачи). Плохая обусловленность является сдерживающим фактором в использовании методов дискриминантного анализа.

В рассматриваемом нами случае, при изучении факторов , влияющих на прогрессирование ИБС, предпосылка дискриминантного анализа отсутствия значимых различий между ковариационными матрицами для исследуемых групп не выполняется.

Подход на основе логистических моделей обладает тем преимуществом, по сравнению с дискриминантным анализом, что обладает гораздо менее жесткими ограничениями, и поэтому имеет более широкую область применения.

Вероятностная природа причинно-следственных связей в развитии ИБС определяет концепцию многофакторности. Научные концепции, основанные на анализе факторов риска в развитии, прогрессировании и исходе патологических состояний и заболеваний имеют практическое подтверждение при изучении заболеваний сердечно-сосудистой системы.

Одним из адекватных методов анализа подобных связей является метод логистической регрессии, который в отечественной биомедицине начал активно применяться в последние 2-3 года. Между тем в зарубежной практике это довольно популярный метод. Так, например, с помощью логистических моделей изучали прогноз коронарной болезни сердца и атеросклероза у взрослого населения [38].

Целью исследования являлось изучение факторов риска сердечно-сосудистых заболеваний, выделение среди них наиболее значимых, оказывающих наиболее значительное влияние на прогрессирование хронической ИБС в виде ишемической делятационной кардиомиопатии, с использованием бинарной логистической регрессии.

Исследование проводилось в два этапа. На первом этапе при помощи бинарной логистической функции моделировалась зависимость наличия или отсутствия ишемической дилятационной кардиомиопатии у пациентов, находящихся на перитонеальном диализе от наличия ряда клинических признаков. На втором этапе производилось тестирование созданной модели с позиций ее статистической значимости.

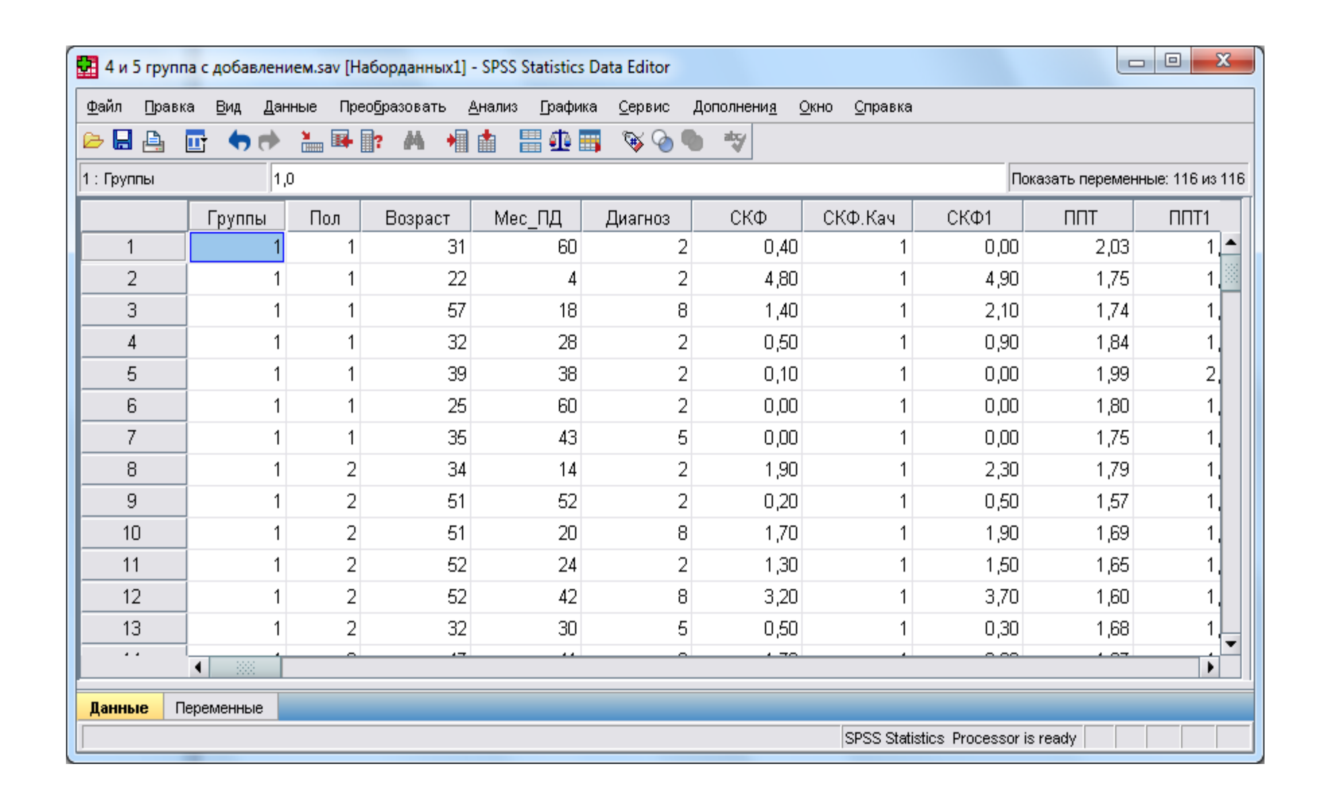
Все пациенты были разделены на группы следующим образом: 1-я группа - пациенты без ишемии миокарда; 2-я группа – пациенты с ишемической дилятационной кардиомиопатией.

В качестве потенциальных предикторов использовали гемодинамические факторы (среднее артериальное давление, гемоглобин, гематокрит и др.); нутриционный статус (альбумин, индекс массы тела, площадь поверхности тела и др); фосфорно-кальциевый обмен (фосфор, кальций, интактный паратиреоидный гормон и др.); дислипидемия (липопротеиды высокой плотности, липопротеиды низкой плотности и др); показатели адекватности перитонеального диализа (средние молекулы, скорость клубочковой фильтрации и др); системное воспаление (интерлекины-1β, интерлейкин-8, С-реактивный белок, сывороточный амилоид А, фибриноген , ферритин, β2- микроглобулин и др.).

Все показатели были закодированы и поставлены соответственно 87-мерному вектору, который учитывает отсутствие, наличие, направленность и величину каждого показателя.

Математическая обработка результатов проводилась с использованием и пакета прикладных программ SPSS 21 для Windows.

Для ввода всех закодированных данных в компьютер, после запуска SPSS для Windows с помощью двойного нажатия левой кнопкой мыши на значке SPSS открыли редактор данных и внесли всю необходимую информацию (рис. 3.1).




| | Группы | Пол | Возраст | Мес_ПД | Диагноз | СКФ | СКФ.Кач | СКФ1 | ППТ | ППТ1 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 31 | 60 | 2 | 0,40 | 1 | 0,00 | 2,03 | 1 |
| 2 | 1 | 1 | 22 | 4 | 2 | 4,80 | 1 | 4,90 | 1,75 | 1 |
| 3 | 1 | 1 | 57 | 18 | 8 | 1,40 | 1 | 2,10 | 1,74 | 1 |
| 4 | 1 | 1 | 32 | 28 | 2 | 0,50 | 1 | 0,90 | 1,84 | 1 |
| 5 | 1 | 1 | 39 | 38 | 2 | 0,10 | 1 | 0,00 | 1,99 | 2 |
| 6 | 1 | 1 | 25 | 60 | 2 | 0,00 | 1 | 0,00 | 1,80 | 1 |
| 7 | 1 | 1 | 35 | 43 | 5 | 0,00 | 1 | 0,00 | 1,75 | 1 |
| 8 | 1 | 2 | 34 | 14 | 2 | 1,90 | 1 | 2,30 | 1,79 | 1 |
| 9 | 1 | 2 | 51 | 52 | 2 | 0,20 | 1 | 0,50 | 1,57 | 1 |
| 10 | 1 | 2 | 51 | 20 | 8 | 1,70 | 1 | 1,90 | 1,69 | 1 |
| 11 | 1 | 2 | 52 | 24 | 2 | 1,30 | 1 | 1,50 | 1,65 | 1 |
| 12 | 1 | 2 | 52 | 42 | 8 | 3,20 | 1 | 3,70 | 1,60 | 1 |
| 13 | 1 | 2 | 32 | 30 | 5 | 0,50 | 1 | 0,30 | 1,68 | 1 |

Рисунок 3.1 – Диалоговое окно SPSS. Внесенная информация

Выбрали в меню «Анализ»→ «Регрессия» → «Логистическая» (рис. 3.2). Открылось диалоговое окно «Логистическая регрессия» (рис. 3.3). Поместили единственную группирующую переменную «группы» (сравниваемые группы), содержащую информацию о принадлежности к одному или второму классу (нет ИБС или ИДК), в поле для зависимых переменных, а все необходимые переменные — в поле ковариат.

В качестве метода использования переменных в вычислениях предварительно бал установлен метод «Принудительное включение», при котором в расчёт одновременно вовлекаются все переменные объявленные ковариатами. Однако количество объявленных ковариат должно быть меньше чем количество наблюдений. Альтернативой здесь являются прямая и обратная селекции (пошаговые методы). Метод прямой селекции начинается с использования одних лишь констант на стартовом этапе, а затем последовательно подключаются переменные, которые демонстрируют сильную корреляцию с зависимыми переменными. Далее опять следует проверка того, какие переменные должны быть исключены, причём в качестве критерия проверки выбирается либо статистика Wald («Включение: Вальд»), либо функция правдоподобия («Включение: ОП»),

либо один из вариантов, называемых «условной статистикой» («Включение: Условное» - которые, однако, не рекомендуются) [39].

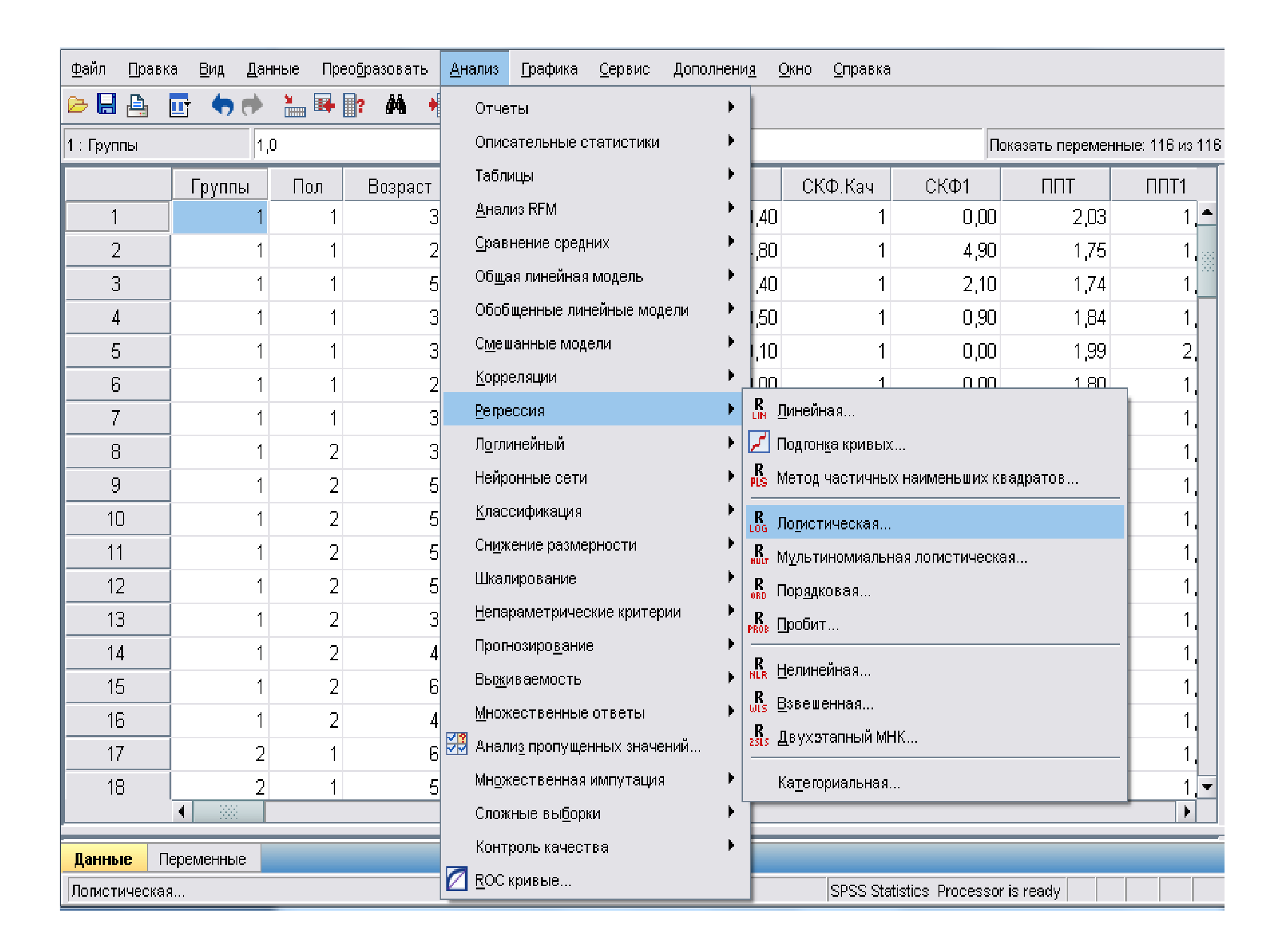



Рисунок 3.2 – Диалоговое окно SPSS. Выбор метода анализа

Метод обратной селекции сначала берёт в расчёт все переменные, а затем в обратном порядке происходит исключение малозначимых переменных. Мы же воспользовались пошаговым методом «Включение ОП» (рис. 3.3), при котором в качестве критерия проверки последовательно включенных переменных выбирается функция правдоподобия.

Используя кнопку «Категориальные» подготовили для расчета категориальные переменные (то есть переменные, принадлежащие к номинальной шкале). На рис. 3.4 представлено диалоговое окно SPSS, которое открывается после нажатия на кнопку «Категориальные» для задания категориальных переменных.

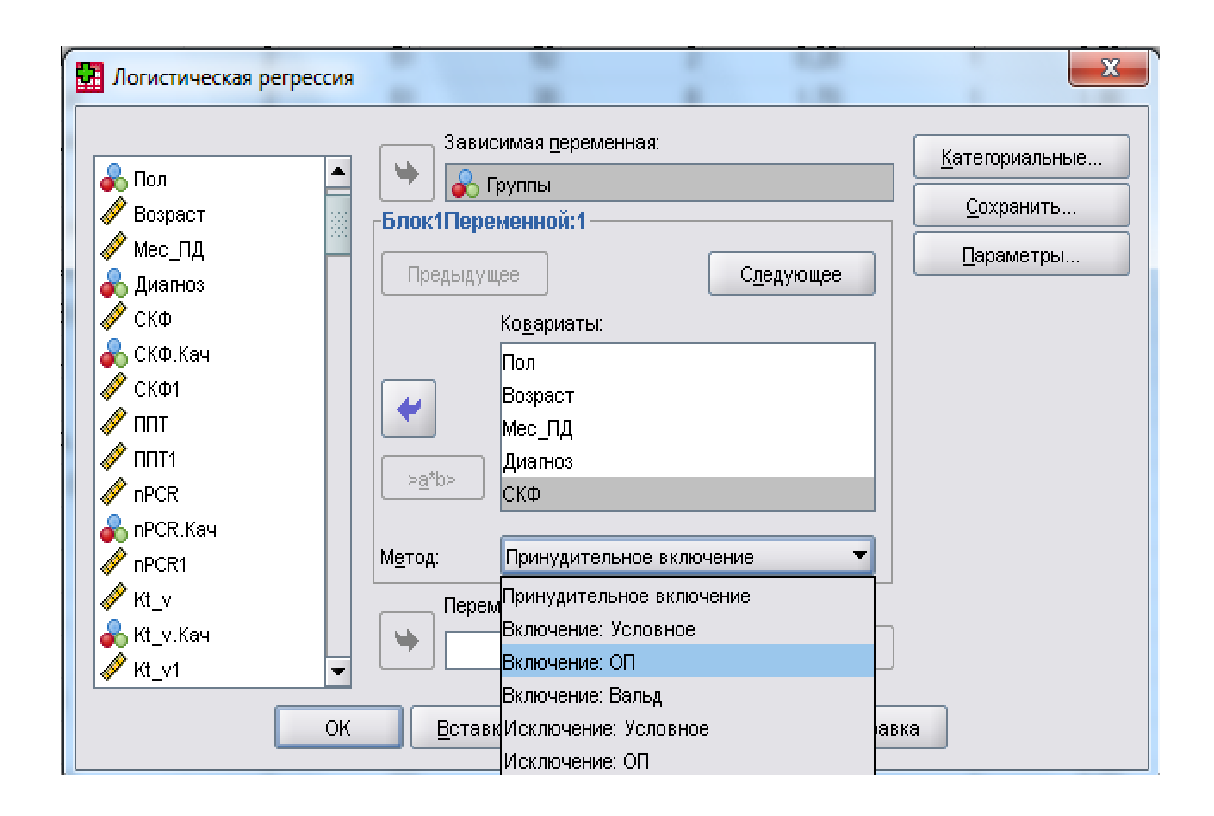



Рисунок 3.3 – Диалоговое окно SPSS. Логистическая регрессия

После установки всех необходимых настроек начали расчёт нажатием кнопки «ОК» (рис. 3.3). Результаты проведенного расчета приведены ниже.

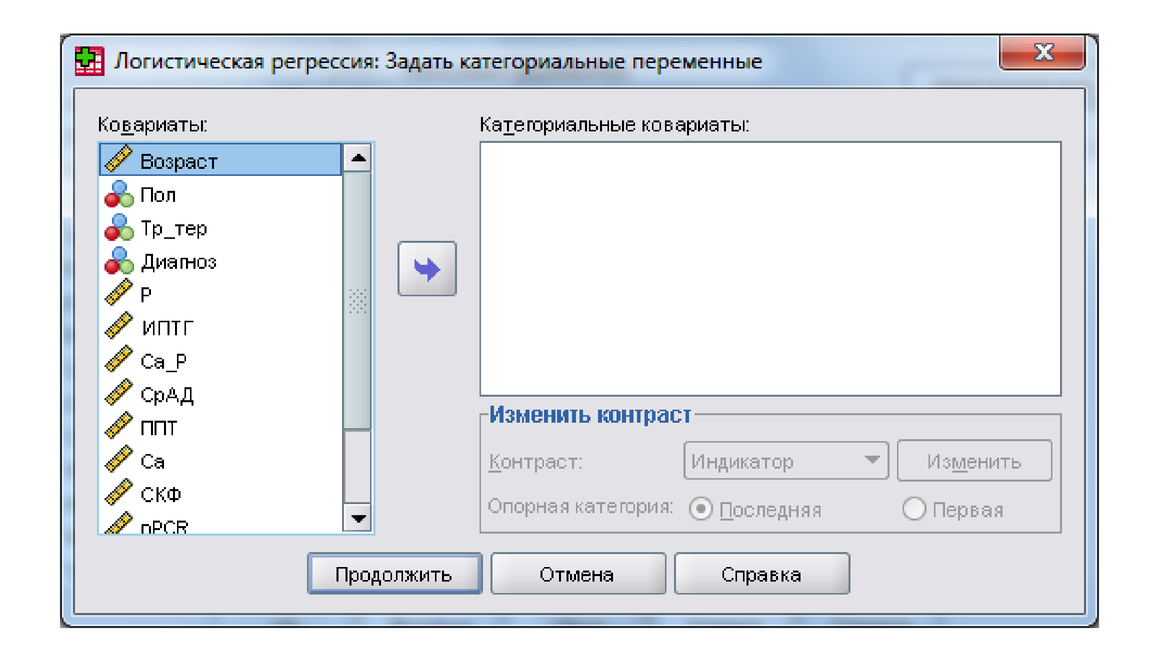



Рисунок 3.4 – Диалоговое окно SPSS. Логистическая регрессия: Задать категориальные переменные

С учетом рассмотренных показателей составили уравнение логистической регрессии, по которому определяется вероятность наличия ишемической дилятационной кардиомиопатии у пациентов, находящихся на перитонеальном диализе:

$$\hat{P} = [1 + \exp(z)]^{-1},$$

где $\hat{P}$ – вероятность того, что произойдет интересующее событие;

z – значение логистической функции, которое определяется согласно формуле:

$$z = b_1 * X_1 + b_2 * X_2 + ... + b_n * X_n + b_0,$$

где $X_i$ – фактор или предиктор,

$b_i$ – коэффициенты, значения которых указывают на изменение отношения шансов при изменении $X_i$.

Отношением шансов является отношение вероятности того, что событие произойдет к вероятности того, что оно не произойдет, высчитывается как P/(1-P).

Параметры модели оценивались с помощью метода максимального правдоподобия [40].

Метод пошагового включения предикторов, который ранжирует признаки в соответствии с их вкладом в модель, применялся при оценке уравнений регрессии. Итогом стало построение регрессионной функции, в которую вошли 4 показателя: $X_1$ - площадь поверхности тела (в м$^2$); $X_2$ - сывороточный амилоид А (в мг/л); $X_3$ - интерлейкин 8 (в пг/мл); $X_4$ - липопротеиды высокой плотности (в ммоль/л).

Подобранная из обучающей выборки бинарная логистическая функция имеет вид:

$$\hat{P} = [1 + \exp(68{,}739 \cdot X_1 + 77{,}496 \cdot X_2 - 0{,}038 \cdot X_3 - 10{,}257 \cdot X_4 - 134{,}818)]^{-1},$$

где $\hat{P}$ - вероятность появления ишемической дилятационной кардиомиопатии у пациентов, находящихся на перитонеальном диализе.

Таким образом, сравнение групп 1 и 2 показало, что из всех оцениваемых факторов статистически значимое влияние на вероятность развития ишемической дилятационной кардиомиопатии у пациентов, находящихся на перитонеальном диализе оказывали липопротеиды высокой плотности, площадь поверхности тела, интерлейкин-8, сывороточный амилоид А.

Проверка значимости подобранных коэффициентов $b_i$ проводилась при помощи статистики Вальда, представляющей собой квадрат отношения соответствующего коэффициента к его стандартной ошибке [41]:

$$W_i = \left( \frac{b_i}{s_i} \right)^2 .$$

Все переменные, согласно статистики Вальда (табл. 3.1), значимы ($p<0,05$) и подобраны правильно.

Рассчитанные коэффициенты регрессионной функции и результаты проверки их значимости представлены в табл. 3.1.

Таблица 3.1

Коэффициенты регрессионной функции

| Показатели ($X_i$) | Коэфи-циенты($b_i$) | Стандартные ошибки ($S_i$) | Критерии Вальда ($W_i$) | Значимость ($P_i$) | Критерий Odds Ratio | 95% Доверительный интервал для Odds Ratio | |
|---|---|---|---|---|---|---|---|
| | | | | | | Нижняя | Верхняя |
| $X_1$ | 68,739 | 28,110 | 5,980 | 0,014 | 7,127E29 | 841956,444 | 6,032E53 |
| $X_2$ | 77,496 | 33,777 | 5,264 | 0,022 | 4,531E33 | 80338,851 | 2,556E62 |
| $X_3$ | -0,038 | 0,016 | 6,020 | 0,014 | 0,963 | ,934 | ,992 |
| $X_4$ | -10,257 | 5,049 | 4,127 | 0,042 | 0,000 | ,000 | ,697 |
| Константа | -134,818 | 55,353 | 5,932 | 0,015 | 0,000 | | |

Полученные результаты свидетельствуют о вероятности перехода от состояния «без ишемии» к «ишемической дилятационной кардиомиопатии».

Столбец $b_i$ показывает изменение вероятности отнесения пациента к группе с ишемической дилятационной кардиомиопатией (в %) при изменении каждого фактора на единицу.

Кроме того, в таблице представлен критерий Odds Ratio - отношение шансов «без ишемии миокарда» и «ишемическая дилятационная кардиомиопатия», который рассчитывается согласно формуле:

$$\text{Odds Ratio} = e^{b_i}.$$

Вывод доверительных интервалов для Odds Ratio осуществляется путем нажатием кнопки «Опции» в диалоговом окне «Логистическая регрессия» (рис. 3.3) и установкой флажка в новом окне (рис. 3.5) напротив поля «ДИ для exp(B)».

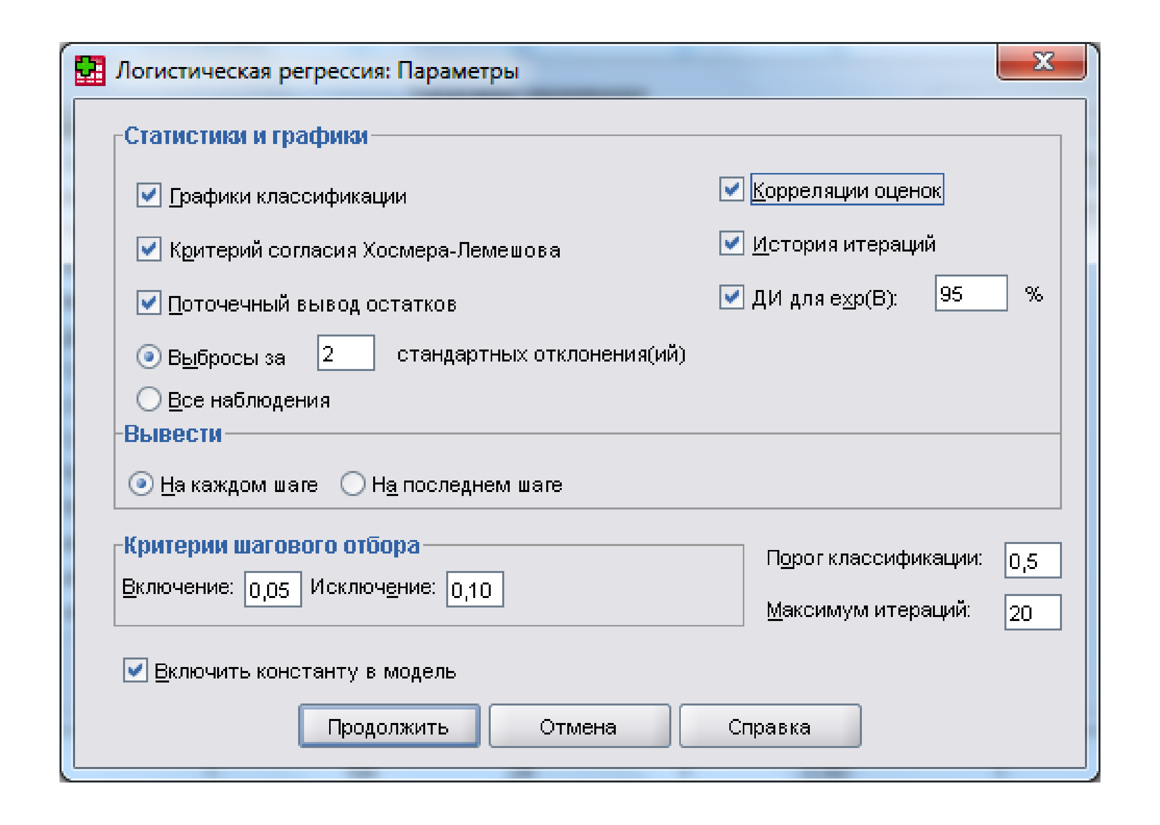



Рисунок 3.5 – Диалоговое окно SPSS.
Логистическая регрессия: Параметры

В области «Критерии шагового отбора» имеются поля «Включение» и «Исключение». В первом указывается значение вероятности, служащее критерием включения предиктора в уравнение регрессии, а во втором — значение вероятности для исключения из уравнения регрессии. По умолчанию в поле «Включение» указано значение 0,05, а в поле «Исключение» — значение 0,10. При желании исследователь может изменять их.

Созданная регрессионная модель имеет высокое качество и для 97,7% объектов результаты прогноза оказались верными (табл. 3.2 и 3.3).

Таблица 3.2

Характеристики модели бинарной логистической регрессии, созданной для определения риска ишемической дилятационной кардиомиопатии у пациентов, находящихся на перитонеальном диализе

| Результаты заключительного шага анализа | -2 Log Правдоподобие (G) | $R^2$ Кокса и Шелла | $R^2$ Нэйджелкерка | $\chi^2$ | P |
|---|---|---|---|---|---|
| | 10,940 | 0,665 | 0,944 | 95,180 | 0,001 |

Таблица 3.3

Классификационные результаты

| Наблюдаемые группы | Спрогнозированные результаты (n) | | % верно предсказанных показателей |
|---|---|---|---|
| | Без ишемии миокарда | Ишемическая дилятационная кардиомиопатия | |
| Без ишемии миокарда | 25 | 1 | 96,2 |
| Ишемическая дилятационная кардиомиопатия | 1 | 60 | 98,4 |
| Суммарный показатель | | | 97,7 |

Качество приближения регрессионной модели оценивается при помощи функции подобия [42]. Мерой правдоподобия служит отрицательное удвоенное значение логарифма этой функции (-2LL). В качестве начального значения для -2LL применяется значение, которое

получается для регрессионной модели, содержащей только константы. Значение -2 Log Правдоподобия определяется как:

$$G = -2(LL_{null} - LL_k),$$

где $LL_{null}$ - логарифм правдоподобия модели, состоящей только из констант [43 – 44].

$LL_{\text{int}ercept}$ - логарифм правдоподобия модели, состоящей из *к* предикторов. При нулевой гипотезе, $H_0$:$b_1$=$b_2$=…=0 статистика G имеет хи-квадрат распределение и оценивает, насколько хорошо независимые факторы влияют на переменную. Полученное в исследовании значение G = 10,940 при P=0,001 показывает, что в целом независимые переменные имеют значительный вклад, чтобы спрогнозировать зависимую переменную [45].

Показатели Кокса и Шелла, Нэйджелкерка, являются мерами определённости. Они указывают на ту часть дисперсии, которую можно объяснить с помощью логистической регрессии. Мера определённости по Коксу и Шеллу имеет тот недостаток, что значение равное 1 является теоретически не достижимым; этот недостаток устранен благодаря модификации данной меры по методу Наделькеркеса. Часть дисперсии, объяснимой с помощью логистической регрессии, в рассматриваемом нами случае составляет 94,4 %.

Формула Кокса и Шелла заключается в следующем [46]:

$$R^2 = 1 - \left(\frac{-2LL_{null}}{-2LL_k}\right)^{\frac{2}{n}}$$

Показатель Нэйджелкерка находится как:

$$R^2 = \frac{1 - \left(\frac{-2LL_{null}}{-2LL_k}\right)^{\frac{2}{n}}}{1 - (-2LL_{null})^{2/n}}$$

На рис. 3.6 представлена диаграмма классификации. В пакете SPSS вывод диаграмм классификации осуществляется установкой флажка в диалоговом окне «Логистическая регрессия: Параметры» (рис. 3.5) напротив поля «Графики классификации».

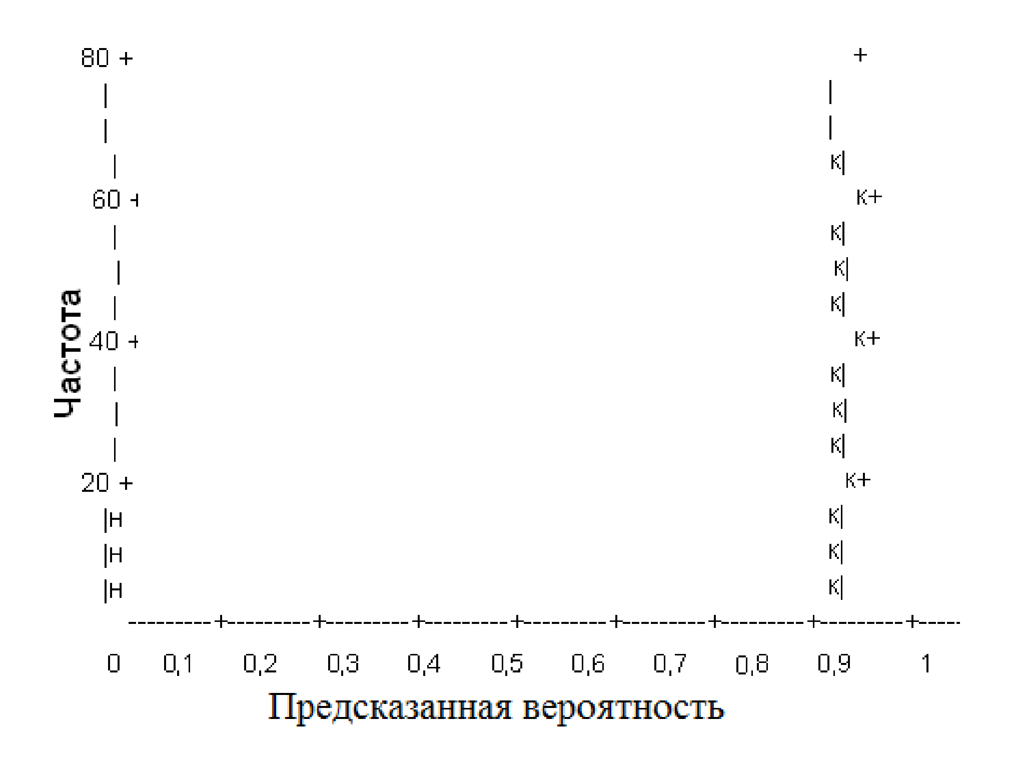



Рисунок 3.6 – Диаграмма классификации.
Символы: н – без ишемии миокарда; к – дилятационная кардиомиопатия

Точкой разделения служило значение р=0,5.

По горизонтальной оси отложены значения предсказанной вероятности, вычисляемые по уравнению регрессии, по вертикали – частоты. Таким образом, каждый столбик на диаграмме соответствует определенной предсказанной вероятности, а его высота — количеству объектов, для которых предсказана данная вероятность. Чем ближе значение прогнозируемой вероятности к единице, тем вероятнее развитие дилятационной кардиомиопатии у пациента, находящегося на перитонеальном диализе.

О количестве правильных и неправильных предсказаний позволяет судить классификационная табл. 3.3.

Из таблицы можно сделать вывод о том, что из общего числа

больных с ишемической дилятационной кардиомиопатией, равного 61, тестом были признаны верно 60 и 1 ошибочно отнесен к группе «Без ишемии миокарда». Из общего числа здоровых («Без ишемии миокарда»), равного 26, тестом были признаны верно 25 и 1 ошибочно отнесен к группе больных. В общем, правильно были распознаны 85 случаев из 87, это составляет 97,7 %.

Общая оценка согласия между влиянием выявленных в модели факторов риска и реально зафиксированным наступлением неблагоприятного исхода производилась с использованием теста согласия Хосмера–Лемешова ($H_L$), в котором значение р тем выше, чем меньше различия между частотой наблюдаемых и предсказанных на основании данных регрессионной модели исходов.

Критерий Хосмера-Лемешова [47 – 48], который показывает степень различия между оцененными и фактическими значениями зависимой переменной в обеих группах (значения бинарной переменной 0 и 1) определяется как:

$$H_L = \sum_{j=1}^{g} \frac{\left(O_j - E_j\right)^2}{E_j \cdot (1 - E_j / n_j)},$$

где $n_j$ - количество наблюдений в $j$-й группе;

$O_j = \sum_i y_{ij} =$ наблюдаемое число случаев в $j$-й группе;

$E_j = \sum_i \hat{p}_{ij} =$ ожидаемое число случаев в $j$-й группе.

Полученное наблюдаемое $O_j$ и ожидаемое $E_j$ число случаев в $j$-й группе для каждого шага представлено в табл. 3.2.

Согласно полученным $O_j$ и $E_j$ (табл. В.1, Приложение В) для каждого шага были рассчитаны значения критерия Хосмера-Лемешова (табл. 3.4). В SPSS расчет теста согласия Хосмера-Лемешова осуществляется путем установки опции «Критерий согласия Хосмера-Лемешова» в диалоговом окне SPSS. Логистическая регрессия: Параметры (рис. 3.5).

Таблица 3.4

Критерий Хосмера-Лемешова

| Шаг | $H_L$ | Степень свободы | Значимость $P_i$ |
|---|---|---|---|
| 1 | 10,066 | 7 | 0,185 |
| 2 | 10,564 | 8 | 0,228 |
| 3 | 0,254 | 5 | 0,998 |
| 4 | 1,227 | 4 | 0,874 |

Полученное на последнем шаге низкое значение $H_L$=1,227, при уровне значимости $P$>0,05 (p=0,874), свидетельствует о приемлемом качестве подобранной модели.

Вывод остатков (рис. 3.7) осуществляется путем установки в диалоговом окне «Логистическая регрессия: Параметры» опции «Поточечный вывод остатков» (рис. 3.5).

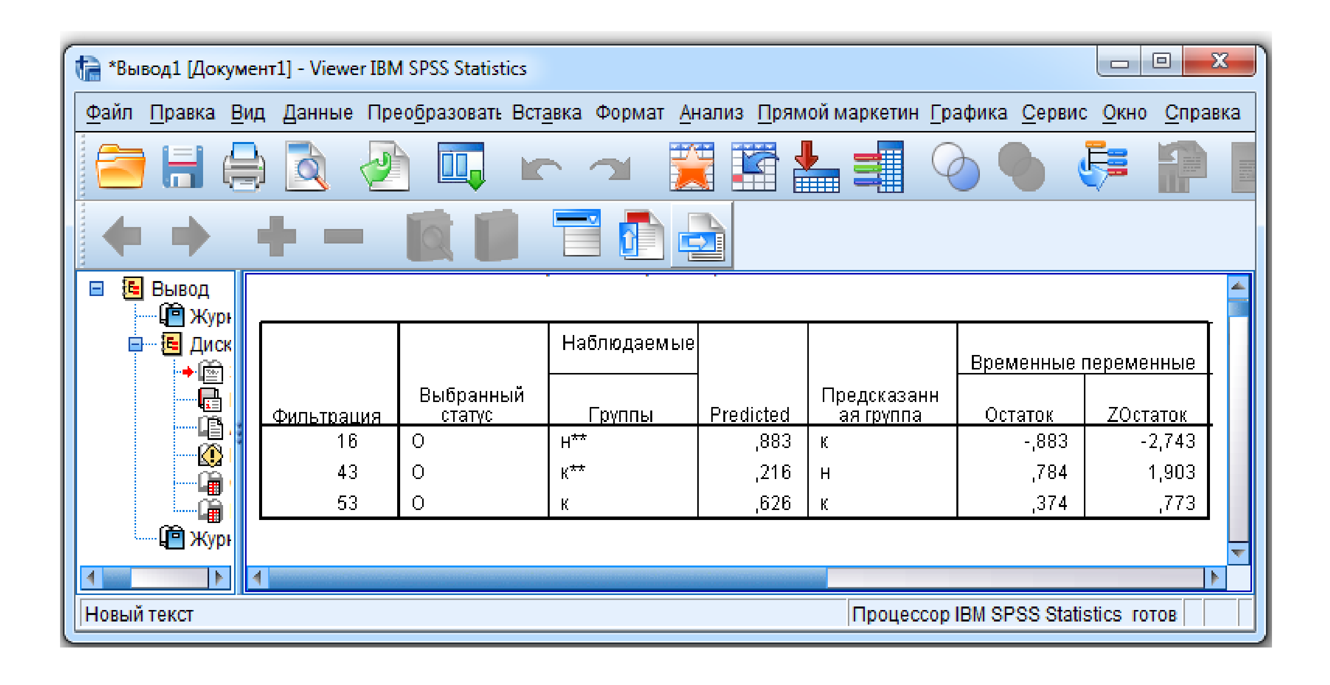
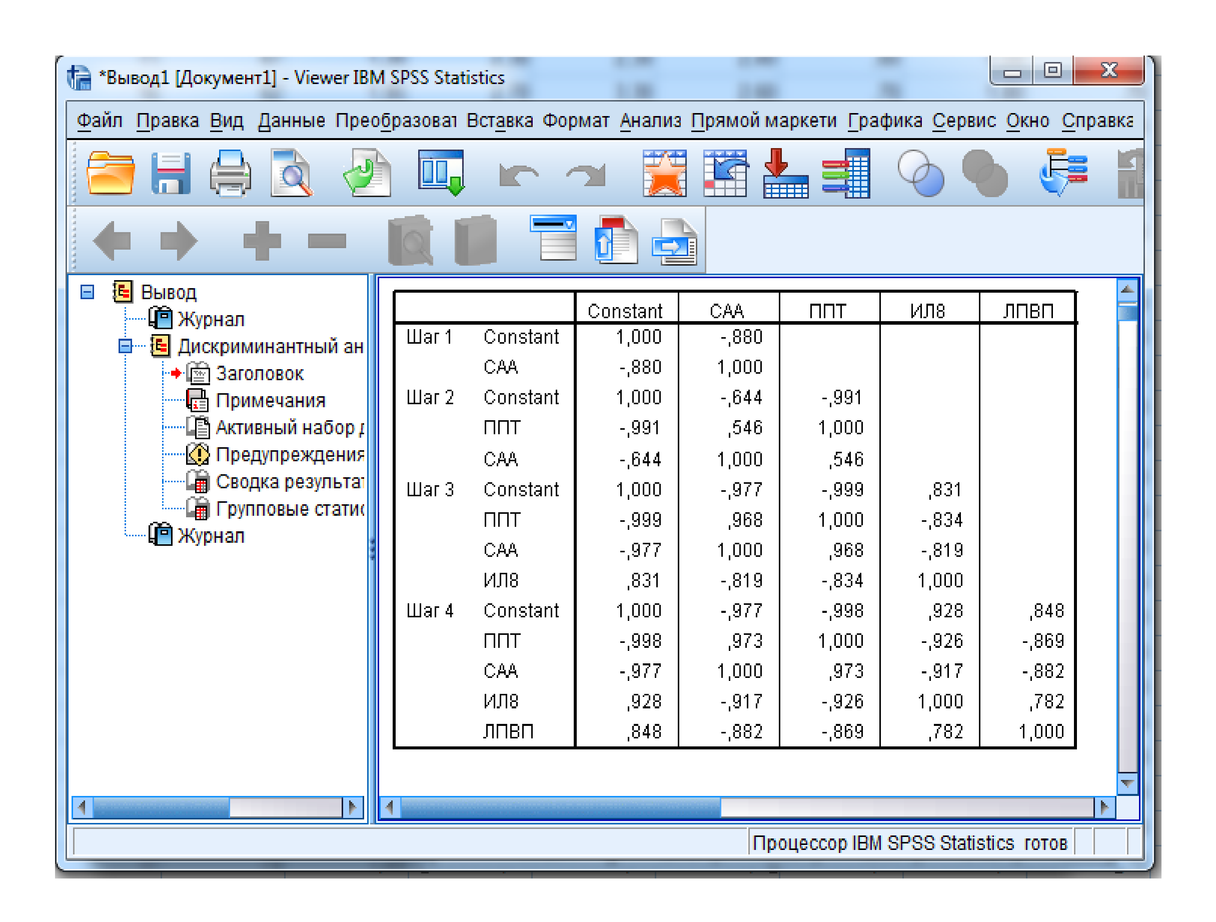



Рисунок 3.7 – Поточечный вывод остатков

В первом столбце таблицы (рис. 3.7) указан номер исследуемого пациента, во втором столбце буквой О обозначены отобранные наблюдения, ** – неверно классифицированные наблюдения. Таким образом, пациент №16 относящийся к группе «Без ишемии миокарда» ошибочно отнесен к группе «Ишемическая дилятационная кардиомиопатия», а пациент под номером №43, относящейся к группе «Ишемическая дилятационная кардиомиопатия», ошибочно отнесен к группе «Без ишемии миокарда».

В пакете SPSS в диалоговом окне «Логистическая регрессия: Параметры» (рис. 3.5) флажок «Корреляции оценок» указывает на необходимость построения корреляционной матрицы для всех переменных, включенных в регрессионное уравнение. Полученная корреляция оценок представлена на рис. 3.8.




| | | Constant | САА | ППТ | ИЛ8 | ЛПВП |
|---|---|---|---|---|---|---|
| Шаг 1 | Constant | 1,000 | -,880 | | | |
| | САА | -,880 | 1,000 | | | |
| Шаг 2 | Constant | 1,000 | -,644 | -,991 | | |
| | ППТ | -,991 | ,546 | 1,000 | | |
| | САА | -,644 | 1,000 | ,546 | | |
| Шаг 3 | Constant | 1,000 | -,977 | -,999 | ,831 | |
| | ППТ | -,999 | ,968 | 1,000 | -,834 | |
| | САА | -,977 | 1,000 | ,968 | -,819 | |
| | ИЛ8 | ,831 | -,819 | -,834 | 1,000 | |
| Шаг 4 | Constant | 1,000 | -,977 | -,998 | ,928 | ,848 |
| | ППТ | -,998 | ,973 | 1,000 | -,926 | -,869 |
| | САА | -,977 | 1,000 | ,973 | -,917 | -,882 |
| | ИЛ8 | ,928 | -,917 | -,926 | 1,000 | ,782 |
| | ЛПВП | ,848 | -,882 | -,869 | ,782 | 1,000 |

Рисунок 3.8 – Корреляция оценок

Также в пакете SPSS есть возможность просмотра истории итераций (рис. В.1, Приложение В) путем установки опции «История итераций» в диалоговом окне «Логистическая регрессия: Параметры» (рис. 3.5).

Для оценки эффективности модели использовался ROC-анализ. Для проведения ROC-анализа на первом этапе получили дополнительные переменные. При помощи кнопки «Сохранить» в диалоговом окне SPSS. «Логистическая регрессия» (рис. 3.3) перешли в окно «Логистическая регрессия: Сохранить» (рис. 3.9) и установили флажки «Вероятности» и «Принадлежность к группе». После чего в файл добавились новые переменные PRE_1 (предсказанные вероятности) и PGR_1 (предсказанные принадлежности в группе).

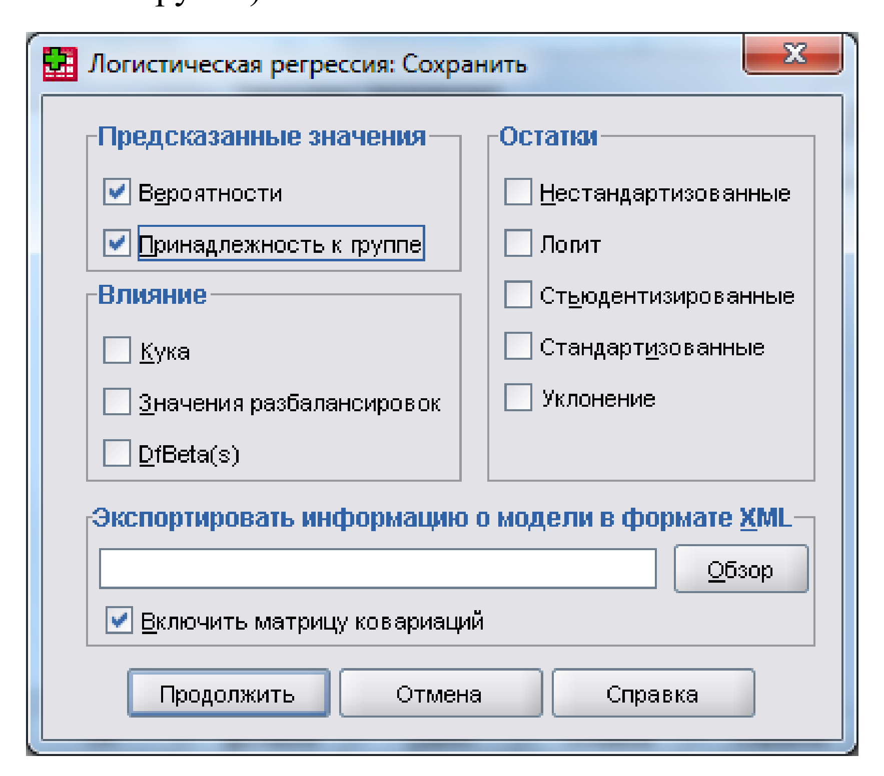



Рисунок 3.9 – Диалоговое окно SPSS. «Логистическая регрессия: Сохранить»

Далее в меню выбрали «Анализ» → «ROC-кривые» как показано на рис. 3.10. В открывшемся диалоговом окне «ROC-кривые» (рис. 3.11) выбрали в качестве проверяемой переменной полученную переменную «PGR_1» (предсказанные принадлежности к группе), а в качестве переменной состояния – переменную «группы» (фактическая принадлежность к группе). Значение переменной состояния установили 2, т.к. 2 – прогнозируемая группа (ишемическая дилятационная кардиомиопатия). После установки всех необходимых настроек нажали кнопку «OK».

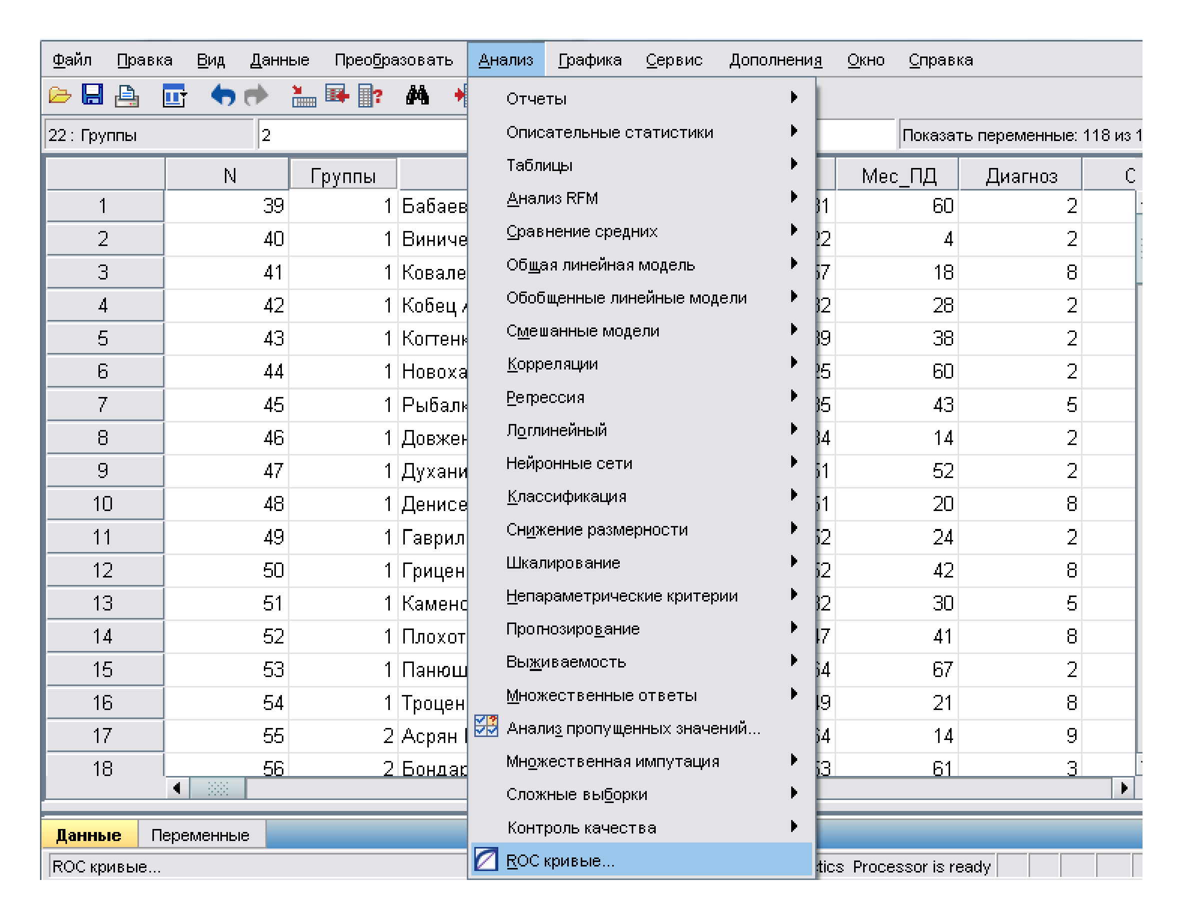



Рисунок 3.10 – Диалоговое окно SPSS. Выбор метода

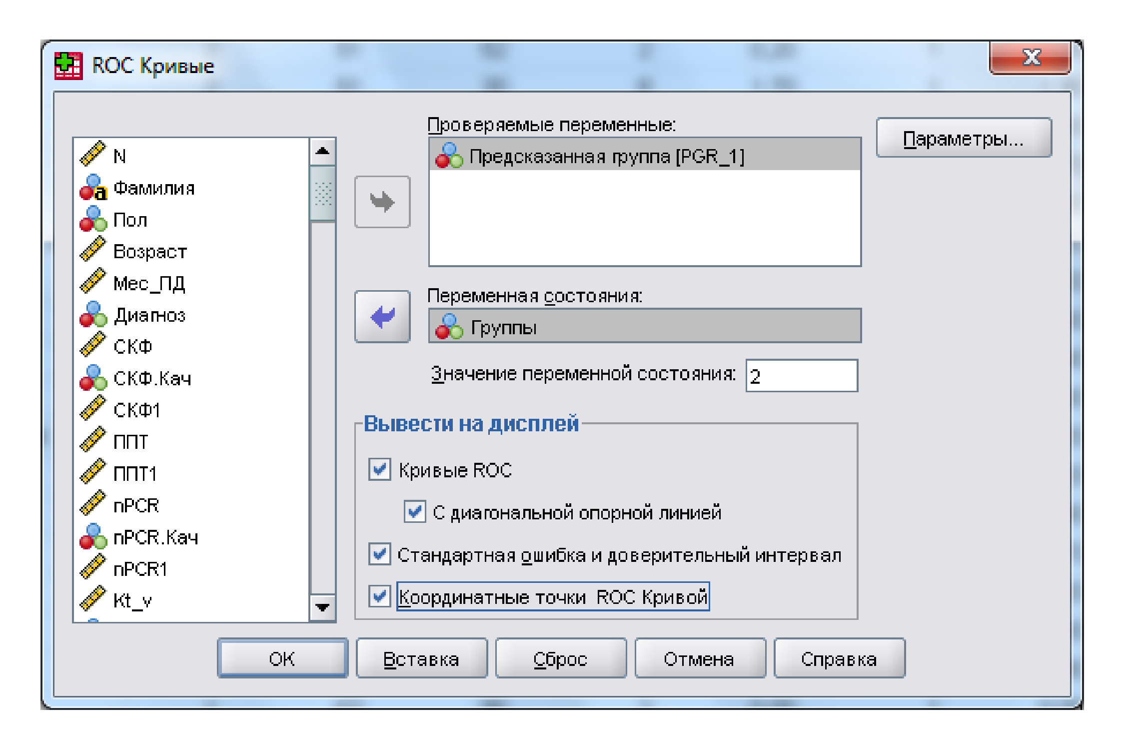



Рисунок 3.11 – Диалоговое окно SPSS. ROC-кривые

Результаты ROC-анализа приведены в табл. 3.5 и на рис. 3.12.

Таблица 3.5

Результаты ROC-анализа

| Характеристики ROC Кривой | | | 95% Доверительный интервал | |
|---|---|---|---|---|
| Площадь | Стандартная ошибка | Значимость *P* | Нижняя граница | Верхняя граница |
| 0,974 | 0,023 | 0,0001 | 0,0001 | 1,000 |

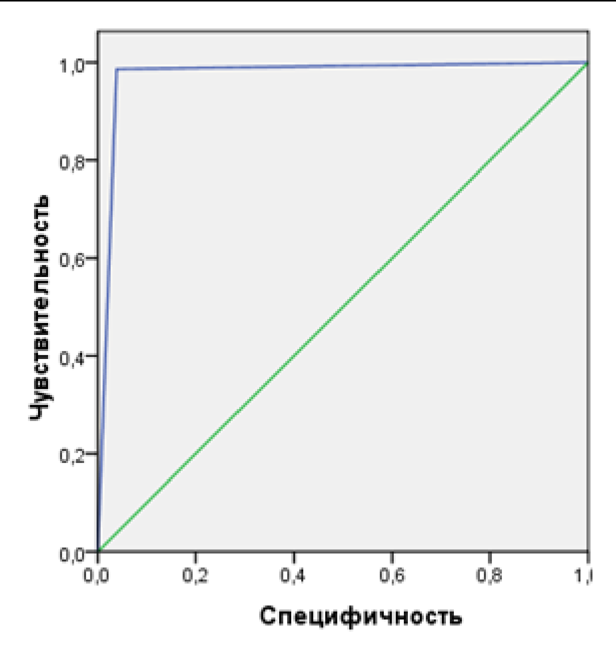



Рисунок 3.12 – ROC Кривая

Значение площади под кривой AUC (Area Under Curve) составило 0,974.

Roc-анализ полученной модели (рис. 3.12, табл. 3.5) выявил ее характеристики, являющиеся показателем отличного качества исследуемой модели.

Таким образом, у пациентов, находящихся на перитонеальном диализе, большая площадь поверхности тела, низкое содержание липопротеидов высокой плотности может служить фактором риска развития и прогрессирования хронической ишемической болезни сердца в виде ишемической дилятационной кардиомиопатии со стойкой эндотелиальной дисфункцией коронарных артерий и диастолической дисфункцией с исходом в диффузный кардиосклероз и сердечную недостаточность.

Применение синтезированной логистической модели позволит проводить своевременную диагностику, лечение и профилактику коронарных событий у больных с почечной недостаточностью.

# ГЛАВА 4. ПРОГНОЗИРОВАНИЕ ЗАБОЛЕВАЕМОСТИ РАССТРОЙСТВАМИ ПСИХИКИ И ПОВЕДЕНИЯ НЕПСИХОТИЧЕСКОГО ХАРАКТЕРА

Из-за социально-экономического кризиса претерпевают быстрые, резкие изменения условия существования людей. Растёт безработица, меняется общественный менталитет, образ жизни, характер питания. Качество жизни и социальное функционирование становятся одним из важнейших показателей психического здоровья. Термин «психическое здоровье», введенный Всемирной организацией здравоохранения, обозначает успешное выполнение психических функций, имеющих результатом продуктивную деятельность, установление отношений с другими людьми и способность адаптироваться к изменениям и справляться с неприятностями. Психическое здоровье – это в первую очередь отсутствие психических расстройств и форм инвалидности. От психических расстройств страдает более 450 миллионов человек. Помимо этого многие люди имеют разнообразные психические проблемы [49 - 50].

Исследования последних лет показали, что у сельских жителей по частоте инвалидизации психические расстройства стоят среди первых трех категорий наряду с сердечно-сосудистыми заболеваниями и травмами.

За период с 2000 по 2010 гг. отмечалось увеличение показателя общей заболеваемости психическими и поведенческими расстройствами у сельского населения Украины. При этом в структуре общей заболеваемости первое ранговое место занимают непсихотические психические расстройства [51].

Весомая доля в структуре психических расстройств и отмечаемый рост заболеваемости непсихотических психических расстройств указывают на актуальность изучения тенденций психического здоровья сельского населения Украины и составления прогнозов заболеваемости, необходимых для планирования ресурсов и объемов медицинской помощи, последующей разработки научно-практического комплекса медико-социальных мероприятий, направленных на оптимизацию психиатрической помощи в Украине, в частности, сельскому населению.

Известно множество методов прогнозирования, но все они имеют одну общую идею: обнаружение связей между прошлым и будущим, между информацией о процессе в контролируемый период времени и

характером протекания процесса в дальнейшем. От того, насколько точно описаны исследуемые связи, будет зависеть точность прогнозирования [52].

В качестве прогнозной модели нами была выбрана регрессионная модель ARMA. ARMA-процессы – это класс стохастических процессов, используемых для анализа временных рядов. Модели позволяют получать точные прогнозы, опираясь только на информацию, содержащуюся в предыстории прогнозируемых рядов медицинских данных.

В практических медицинских задачах можно было бы обойтись без использования моделей ARMA, ограничиваясь либо AR либо MA моделями. Однако, при таком подходе, количество коэффициентов, подлежащих оцениванию, может оказаться слишком большим (что снижает точность оценивания) и даже превосходить количество имеющихся наблюдений. В этом смысле модели ARMA являются «более экономными».

Анализ научных публикации последнего десятилетия показал, что в большом количестве работ по прогнозированию классическими методами используется именно ARMA, как один из наиболее обоснованных и надежных алгоритмов (из статистических).

Прогнозирование заболеваемости непсихотическими психическими расстройствами сельского населения различных регионов Украины позволит усовершенствовать систему оказания психиатрической помощи сельскому населению Украины.

Метод прогнозирования заболеваемости сельского населения Украины непсихотическими психическими расстройствами включает в себя ряд этапов.

**Этап 1.** Сбор статистической информации, характеризующей заболеваемость непсихотическими психическими расстройствами сельского населения Украины за несколько лет с шагом в полгода.

Для построения модели прогнозирования заболеваемости расстройствами психики и поведения непсихотического характера среди сельского населения Украины, проживающего в 24 областях страны, использовали статистические данные за 2008 -2012 гг. Исходные данные сформировали согласно регионального принципа разделения: первый регион - западноукраинский (Волынская, Ровенская, Львовская, Тернопольская, Ивано - Франковская, Закарпатская, Черновицкая области); второй регион - центрально-южный (Киевская, Винницкая,

Кировоградская, Одесская, Николаевская, Херсонская области и АР Крым); третий регион - центрально-северный (Житомирская, Хмельницкая, Черкасская, Полтавская, Сумская, Черниговская области); четвертые регион - восточный (Харьковская, Донецкая, Луганская, Днепропетровская, Запорожская области).

**Этап 2.** Предварительный анализ данных.

На данном этапе проводится структурирование информации, расчет показателей, характеризующих изменение временного ряда, выявление аномальных отклонений.

Сформированные данные о количестве случаев заболеваемости расстройствами психики и поведения представляют собой временные ряды (табл. 4.1).

Таблица 4.1

Заболеваемость расстройствами психики и поведения непсихотического характера сельского населения в 4-х регионах Украины за 2008-2012 года

| Период | Регионы (количество человек) | | | |
|---|---|---|---|---|
| | 1-й регион | 2-й регион | 3-й регион | 4-й регион |
| с 01.01.2008 по 30.06.2008 | 6585 | 7222 | 5305 | 2588 |
| с 01.07.2008 по 31.12.2008 | 6362 | 6891 | 5302,5 | 2625,5 |
| с 01.01.2009 по по 30.06.2009 | 6139 | 6560 | 5300 | 2663 |
| с 01.07.2009 по 31.12.2009 | 6303,5 | 6685,5 | 5151 | 2604,5 |
| с 01.01.2010 по 30.06.2010 | 6468 | 6811 | 5002 | 2546 |
| с 01.07.2010 по 31.12.2010 | 7398,5 | 7784,5 | 6170,5 | 3268 |
| с 01.01.2011 по по 30.06.2011 | 8329 | 8758 | 7339 | 3990 |
| с 01.07.2011 по 31.12.2011 | 8161 | 8720,5 | 7209 | 3935 |
| с 01.01.2012 по по 30.06.2012 | 7993 | 8683 | 7079 | 3880 |
| с 01.07.2012 по 31.12.2012 | 7825 | 8646 | 6949 | 3825 |

По данным, представленным на графиках (рис. 4.1), можно сделать вывод, что заболеваемость непсихотическим психическими расстройствами у сельского населения на протяжении 2011 - 2012 г.г. носила отрицательный характер, то есть наметился спад уровня заболеваемости по всем регионам Украины , что коррелирует с изменением уровня инфляции в этот период.

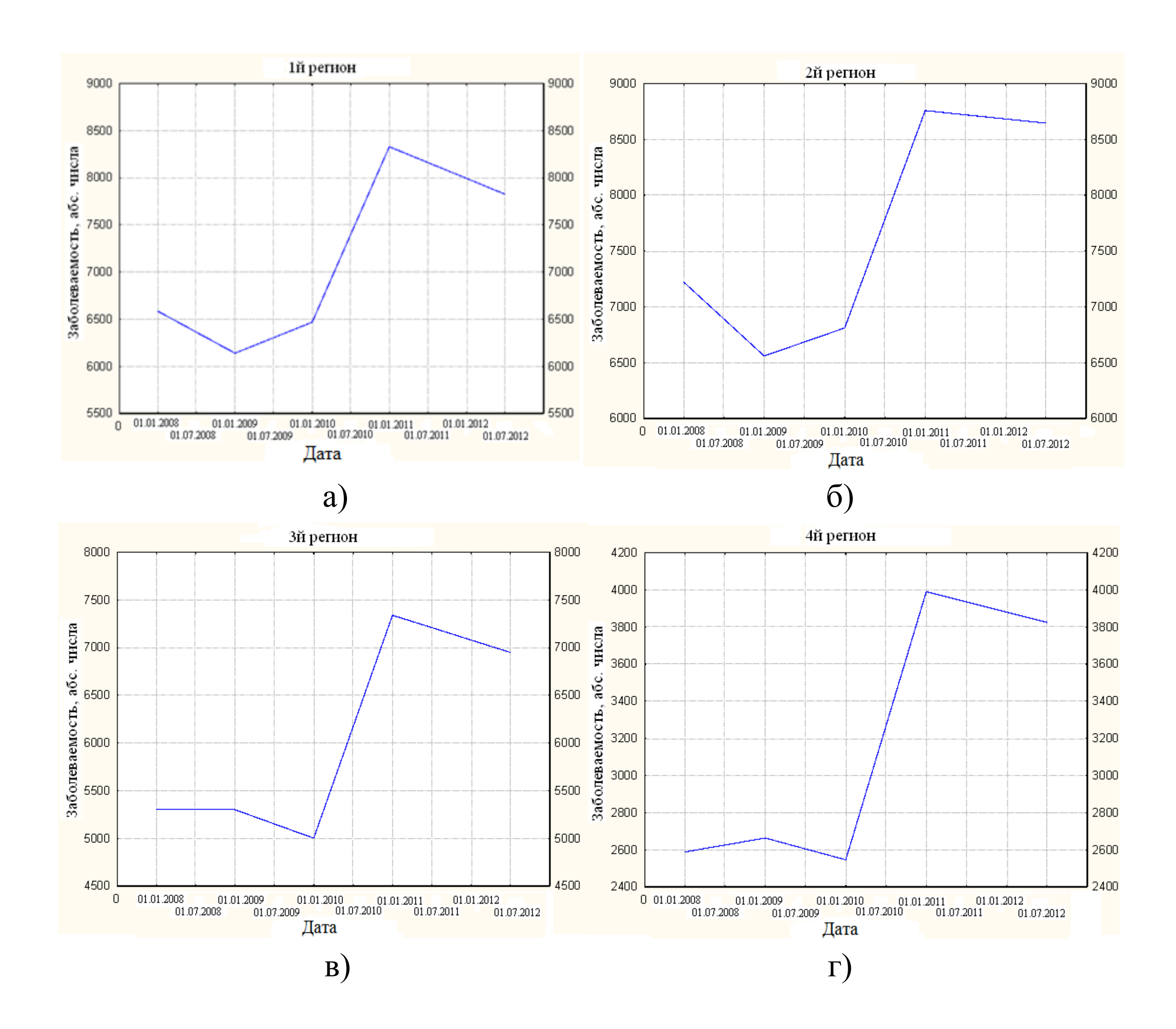



Рисунок 4.1 – Распределение заболеваемости непсихотическими психическими расстройствами сельского населения за 2008-2012 г. по 4-м регионам Украины: а) – по 1-му региону; б) – по 2-му региону; в) – по 3-му региону; г) – по 4-му региону

Рост заболеваемости в 2008 - 2011 гг. и снижение в 2012 г. повторяет тенденции, характерные для страны в целом, а именно: низкий уровень

жизни, высокий уровень безработицы, высокая миграционная активность населения, низкий уровень рождаемости.

Рассмотрим этот факт подробнее.

Большой вред Украине нанес мировой финансовый кризис.

На рис 4.2. представлен график изменения индекса инфляции в Украине, начиная с 2001 года нарастающим итогом.

Согласно данным Государственного комитета статистики Украины [53] количество официально зарегистрированных безработных в 2011 году увеличилось на 5,3% по сравнению с 2010 годом.

Однако в 2012 году безработица в Украине постепенно начала снижаться. Уровень зарегистрированной безработицы в целом по Украине в конце 2011 года составил 1,5% от общего количества трудоспособного населения, что на 0, 7 % меньше, чем 1 марта 2011 года.

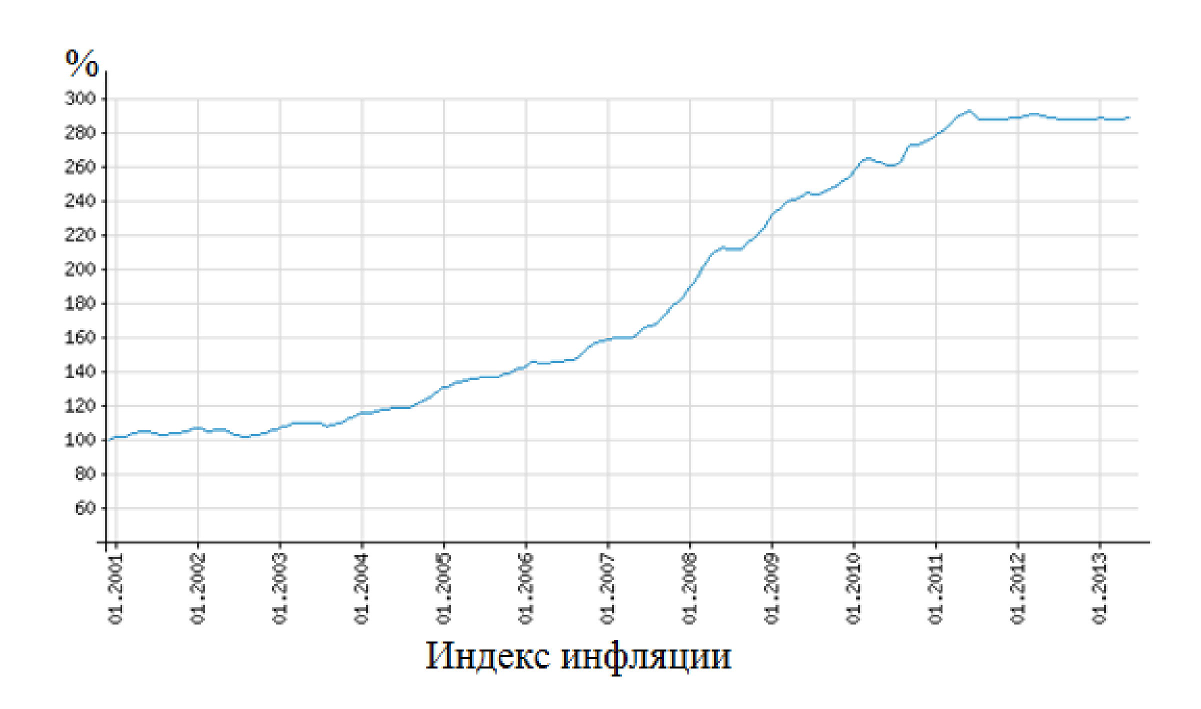



Рисунок 4.2 – График изменения индекса инфляции в Украине

При этом, следует отметить, что 2 миллиона сельских жителей живут только за счет приусадебного хозяйства, многие находятся в условиях неполной занятости.

На этапе предварительного анализа данных проводилось выявление аномальных отклонений с помощью критерия Ирвина:

$$\lambda_t = \frac{|y_t - y_{t-1}|}{S},$$

$$S = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}\left(y_t - \bar{y}\right)^2},$$

где S – несмещённое среднеквадратическое отклонение данного ряда,

$\bar{y}$ – выборочное среднее ряда или средний уровень ряда.

Далее проверили гипотезу $H_0$: «аномальные данные отсутствуют». По таблицам для критерия Ирвина определили $\lambda_{критическое}$.

Если $\lambda_t < \lambda_{критическое}$, то $y_t$ считается нормальным;

если $\lambda_t > \lambda_{критическое}$, то $y_t$ считается аномальным.

Результаты расчетов значений критерия Ирвина для всех 4-х регионов Украины представлены в табл. 4.2., критические значения этого критерия – в табл. 4.3.

Таблица 4.2

Результаты расчетов значений критерия Ирвина для 4-х регионов Украины

| Наблюдение, $n$ | Критерий Ирвина, $\lambda_t$ | | | |
|---|---|---|---|---|
| | 1-й регион | 2-й регион | 3-й регион | 4-й регион |
| 1 | 0,9201553854 | 0,9408362265 | 0,8724324502 | 0,8106499608 |
| 2 | 0,2568567858 | 0,3505195001 | 0,0025785949 | 0,0577362738 |
| 3 | 0,2568567858 | 0,3505195001 | 0,0025785949 | 0,0577362738 |
| 4 | 0,1894750729 | 0,1329008981 | 0,1536842581 | 0,0900685872 |
| 5 | 0,1894750729 | 0,1329008981 | 0,1536842581 | 0,0900685872 |
| 6 | 1,0717723729 | 1,0309085599 | 1,2052352723 | 1,1116157258 |
| 7 | 1,0717723729 | 1,0309085599 | 1,2052352723 | 1,1116157258 |
| 8 | 0,1935064574 | 0,0397114237 | 0,1340869366 | 0,0846798683 |
| 9 | 0,1935064574 | 0,0397114237 | 0,1340869366 | 0,0846798683 |
| 10 | 0,1935064574 | 0,0391819381 | 0,1340869366 | 0,0846798683 |

Таблица 4.3

Критические значения критерия Ирвина

| Число наблюдений n | $\lambda_{критическое}$ | |
|---|---|---|
| | P=0,95 | P=0,99 |
| 2 | 2,8 | 3,7 |
| 3 | 2,2 | 2,9 |
| 10 | 1,5 | 2,0 |
| 20 | 1,3 | 1,8 |
| 30 | 1,2 | 1,7 |
| 50 | 1,1 | 1,6 |
| 100 | 1,0 | 1,5 |
| 400 | 0,9 | 1,3 |
| 1000 | 0,8 | 1,2 |

Согласно табл. 4.3 получили $\lambda_t < \lambda_{критическое}$. Таким образом, в рассматриваемых нами случаях аномальные данные отсутствуют.

Для подтверждения стационарности исследуемых временных рядов провели расширенный тест Дики-Фуллера (ADF-тест), где нулевой гипотезой ($H_{0)}$ являлся тот факт, что исследуемый временной ряд нестационарен и описывается моделями авторегрессии первого порядка. После чего вычисляли значение t-статистики для проверки нулевой гипотезы: $t = \frac{\Delta y_{\mathrm{t}} - 1}{y_t - \Delta \mathrm{y}_{\mathrm{t}}} \cdot (n - L)$, где n – число наблюдений, $L$ – оператор лагов (задержки). Полученное значение сравнивали с критическим ($t_{крит}$). Гипотеза о нестационарности временного ряда отвергали, если $t < t_{крит}$.

В табл. 4.4 приведены значения статистики Дикки-Фуллера ($t_{\varphi}$), а также критические значения ($t_{крит}$) этой статистики на 1%-м, 5%-м и 10%-м уровнях значимости для 1-го региона Украины.

Как видно из таблицы результатов ADF-теста временной ряд для первого региона можно признать стационарными, так как критическое значенея при уровне значимости α = 0,01 превышает расчетное значение статистики Дикки-Фуллера ($t_{\varphi} < t_{крит}$).

Таблица 4.4

Результаты теста Дикки-Фуллера для 1-го региона

| Статистика Дикки-Фуллера ($t_{\varphi}$) | | t-Статистика | Значимость* |
|---|---|---|---|
| | | 1.558967 | 0.4561 |
| Критические значения ($t_{крит}$): | 1% уровень | 4.582648 | |
| | 5% уровень | 3.320969 | |
| | 10% уровень | 2.801384 | |

Анализ результатов ADF-теста для Центрально-Южного, Центрально-Северного и Восточного регионов Украины также подтвердил предположение о стационарности исследуемых временных рядов.

Средние и дисперсию временных рядов рассчитывали по формулам [54]:

$$\bar{y} = \sum_{t=1}^{n} y_t \quad \text{и} \quad s^2 = \frac{1}{n}\sum_{t=1}^{n}\left(y_t - \bar{y}\right)^2 .$$

Выборочную автоковариацию *k*-го порядка вычисляли как:

$$c_k = \frac{1}{n}\sum_{t=1}^{n-k}\left(y_t - \bar{y}\right)\left(y_{t+k} - \bar{\bar{x}}\right) .$$

Количественно автокорреляция устанавливается с помощью линейного коэффициента корреляции между уровнями изучаемого ряда и некоторым лагом $\tau$=k. Величина лага определяет порядок коэффициента корреляции. С увеличением лага число пар уровней, по которым рассчитывается коэффициент автокорреляции, уменьшается. Считается целесообразным использовать коэффициенты автокорреляции с порядками: $k \leq n-4$.

Расчет коэффициентов АКФ *k*-го порядка проводится согласно формулы [55]:

$$r_k = \frac{\sum_{t=k+1}^{n}(y_t - \overline{y_{2k-1}})(y_{t-k} - \overline{y_{2k}})}{\sqrt{\sum_{t=k+1}^{n}(y_t - \overline{y_{2k-1}})^2 \sum_{t=k+1}^{n}(y_{t-k} - \overline{y_{2k}})^2}},$$

где $\overline{y_{2k-1}} = \frac{1}{n-k} \bullet \sum_{t=k+1}^{n} y_t$, $\overline{y_{2k}} = \frac{1}{n-k} \bullet \sum_{t=k+1}^{n} y_{t-k}$, $y_t$– значения временного ряда, *t, k, n*– порядки взятых значений.

Коэффициент автокорреляции первого порядка можно рассчитать как:

$$r_1 = \frac{\mathrm{cov}(y_{t-1}; y_t)}{\sigma_{y_{t-1}} \cdot \sigma_{y_t}} = \frac{\overline{y_{t-1} \cdot y_t} - \overline{y_{t-1}} \cdot \overline{y_t}}{\sqrt{\overline{y_{t-1}^2} - \overline{y_{t-1}}^2} \cdot \sqrt{\overline{y_t^2} - \overline{y_t}^2}},$$

где $\bar{y}_t = \frac{1}{n-1}\sum_{t-2}^{n} y_t, \bar{y}_{t-1} = \frac{1}{n-1}\sum_{t-1}^{n-1} y_t$.

Для проверки гипотезы о том, что наблюдаемые данные являются реализацией «белого шума», использовали Q-статистику Льюинга – Бокса, которую рассчитывали по формуле [56]:

$$\tilde{Q} = n(n+2)\sum_{k=1}^{m}\frac{r_k^2}{n-k}, \quad (4.1)$$

где $r_k$ - автокорреляция k-го порядка,

m - число проверяемых лагов;

n – число наблюдений (n=10).

В предположении, что ряд представляет собой белый шум, использовали следующий доверительный интервал для отдельного коэффициента автокорреляции:

$$\left[ r_k - \sqrt{\frac{n-k}{n(n+2)}\chi^2_{1-\alpha,m}},\ r_k + \sqrt{\frac{n-k}{n(n+2)}\chi^2_{1-\alpha,m}} \right],$$

где $\chi^2_{1-\alpha,m}$ — квантиль распределения $\chi^2$ с m-степенями свободы и уровнем значимости α.

Нулевая гипотеза в $Q$-критерии заключается в том, что ряд представляет собой белый шум, то есть является чисто случайным процессом. Используется стандартная процедура проверки: если расчетное значение $Q$-статистики больше заданного квантиля распределения $\chi^2_{1-\alpha,m}$, то нулевая гипотеза отвергается и признается наличие автокорреляции до $m$-го порядка в исследуемом ряде.

**Этап 3.** Определение параметров моделей (параметрическая идентификация).

Так как исследуемые ряды являются стационарными, то для прогноза заболеваемости психическими расстройствами непсихотического типа сельского населения Украины была выбрана модель Бокса-Дженкинса *ARMA*($p$, $q$), которая обобщает стационарные линейные модели авторегрессии и скользящего среднего ($p$ – параметры авторегрессии, $q$ – параметры скользящего среднего).

Таким образом, временной ряд представили в виде суммы двух процессов: авторегрессии порядка p, и скользящего среднего порядка q:

$$y_t = \varphi_1 y_{t-1} + \ldots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} + \varepsilon_t,$$

где $y_t$ – зависимая переменная в момент времени $t$;

$\varphi_1, \varphi_2, \ldots \varphi_k$ – оцениваемые параметры (параметры авторегрессии);

$\theta_i$ – коэффициенты процесса скользящего среднего,

$\varepsilon_t$ – случайная величина.

Задача заключалась в выборе наилучшего сочетания порядка авторегрессии p и порядка скользящего среднего q.

При попытке построить лучшую модель испытывались разные порядки авторегрессии и скользящего среднего.

Построение моделей прогнозирования осуществили с помощью пакета прикладных программ SPSS 21. В редакторе данных ввели информацию о количестве случаев заболеваемости расстройствами психики и поведения непсихотического характера сельского населения в 4-х регионах Украины (рис. 4.3).

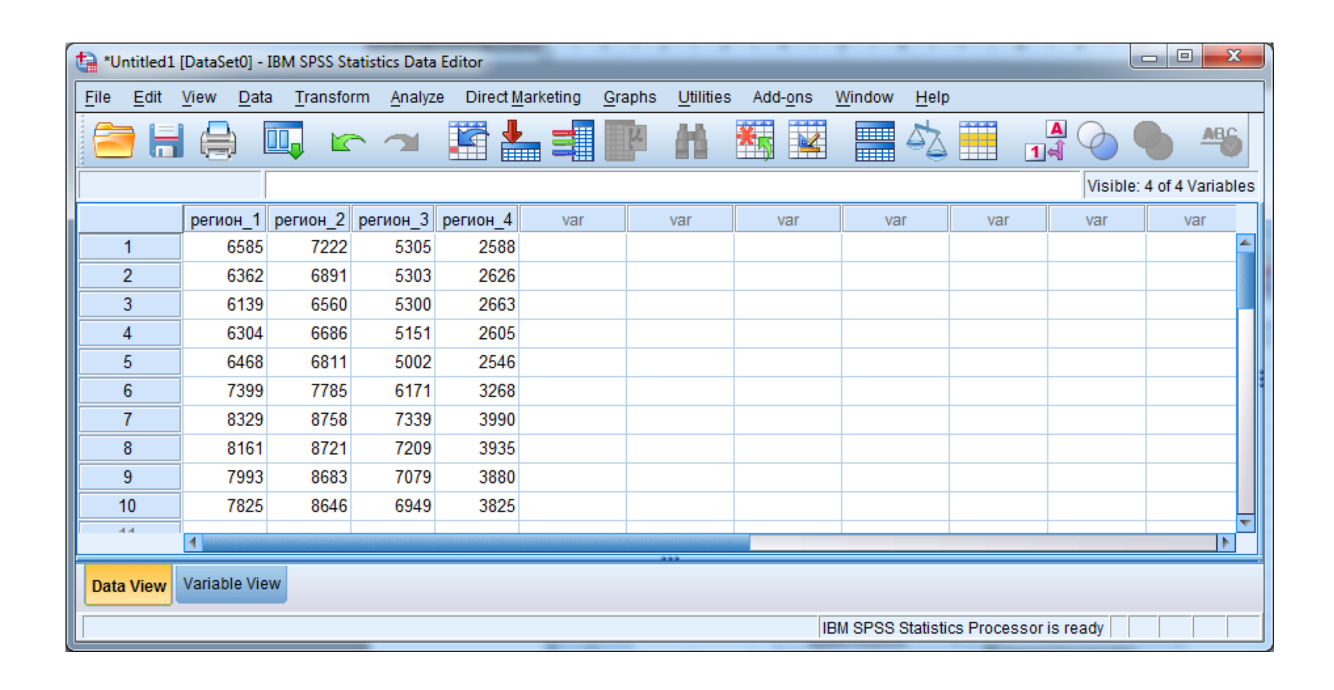



Рисунок 4.3 – Диалоговое окно SPSS

В меню программы выбрали Analyze→Forecasting→Create Models…(рис. 4.4).

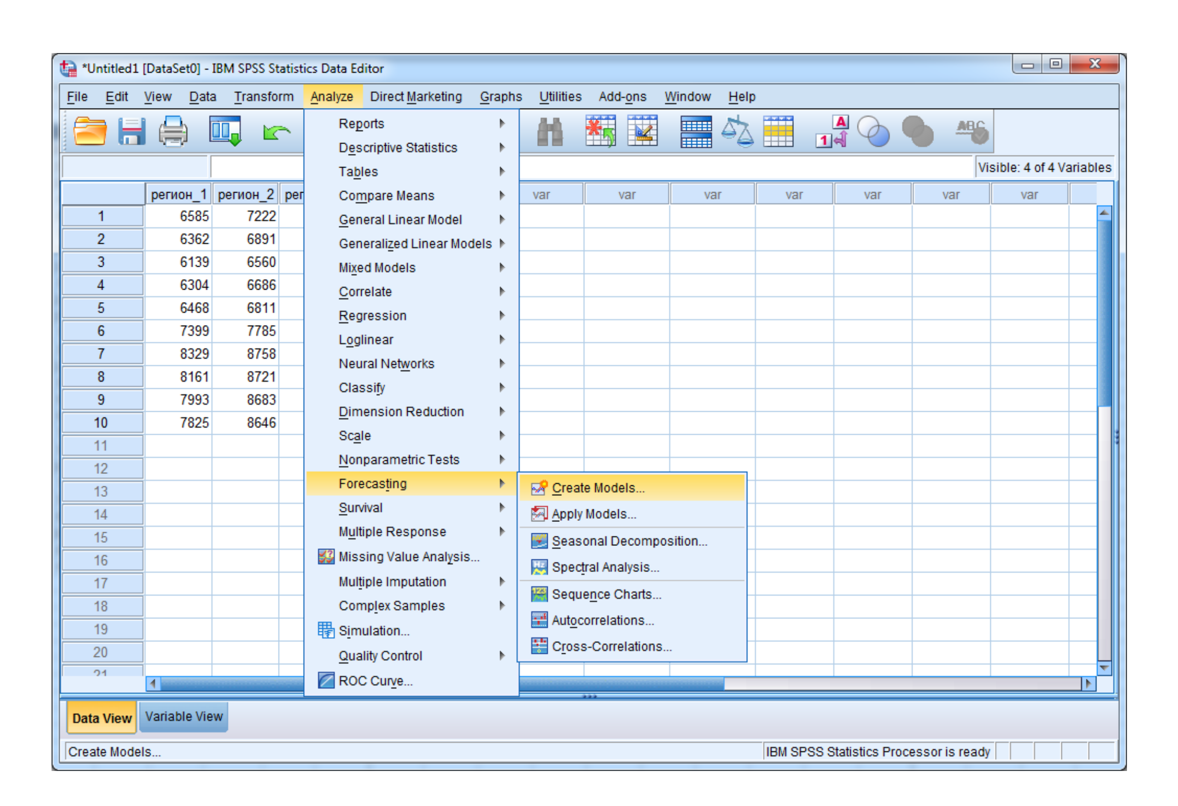



Рисунок 4.4 – Выбор метода прогнозирования

В появившемся окне «Time Series Modeler» (рис. 4.5) в поле «Dependent Variables» внесли зависимую переменную «регион_1». Для

получения моделей по остальным 3-м регионам соответственно внесли в качестве зависимых переменных «регион_2» или «регион_3» или «регион_4». В поле «Method» выбрали метод «ARIMA» и нажали кнопку «Criteria…» (рис.4.6).

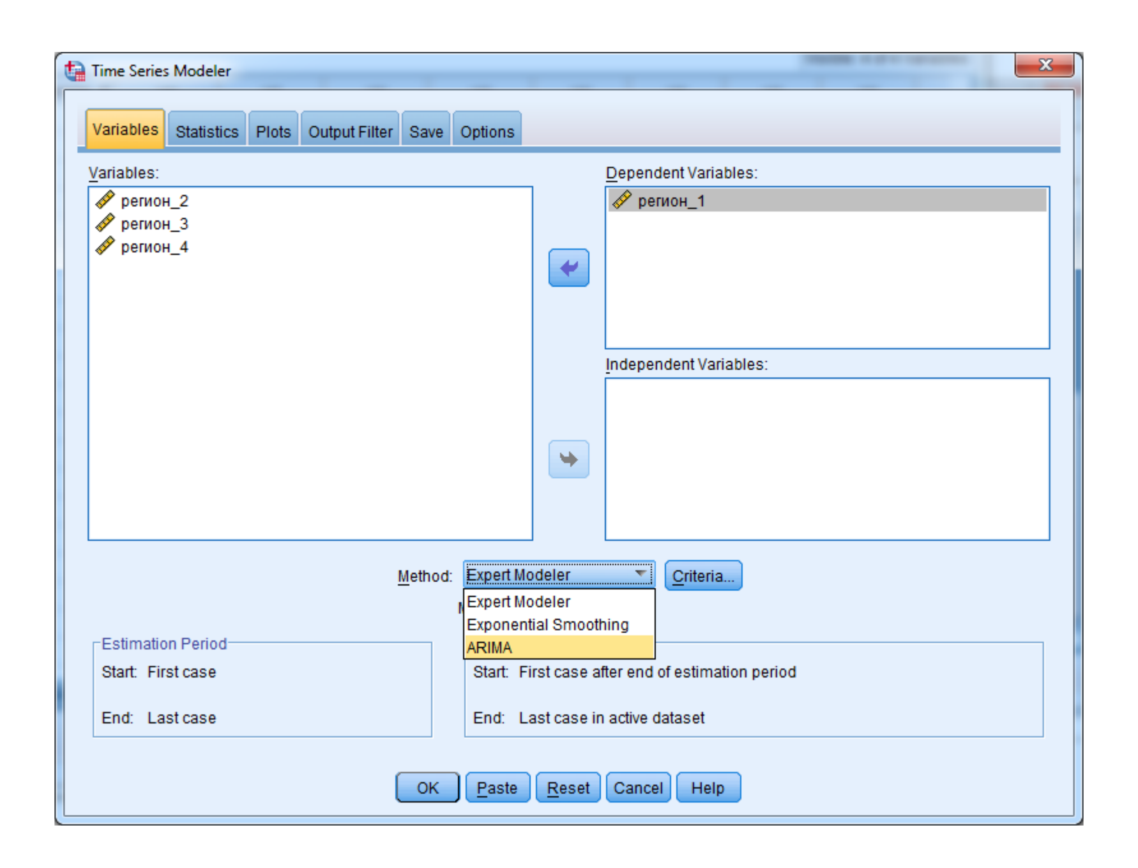



Рисунок 4.5 – Диалоговое окно «Time Series Modeler»

В открывшемся окне «Time Series Modeler: ARIMA Criteria» (рис. 4.6) задали параметры модели ARMA *p* и *q*. Для этого в поле «Nonseasonal» напротив «Autoregressive (p)» двойным нажатием графического манипулятора ввели значение параметра $p$=2 напротив «Moving Average (q)» - $q$=1. Для построения моделей Бокса-Дженкинса ARMA (1, 2) и ARMA (2, 2) аналогичным образом задали параметры *p* и *q*. Далее нажали кнопку «Continue».

Для получения расчетных коэффициентов моделей ARMA для первого региона в диалоговом окне Time Series Modeler» (рис. 4.5) перешли на вкладку «Statistics» (рис. 4.7) и напротив опции «Parameter estimates» установили «галочку».

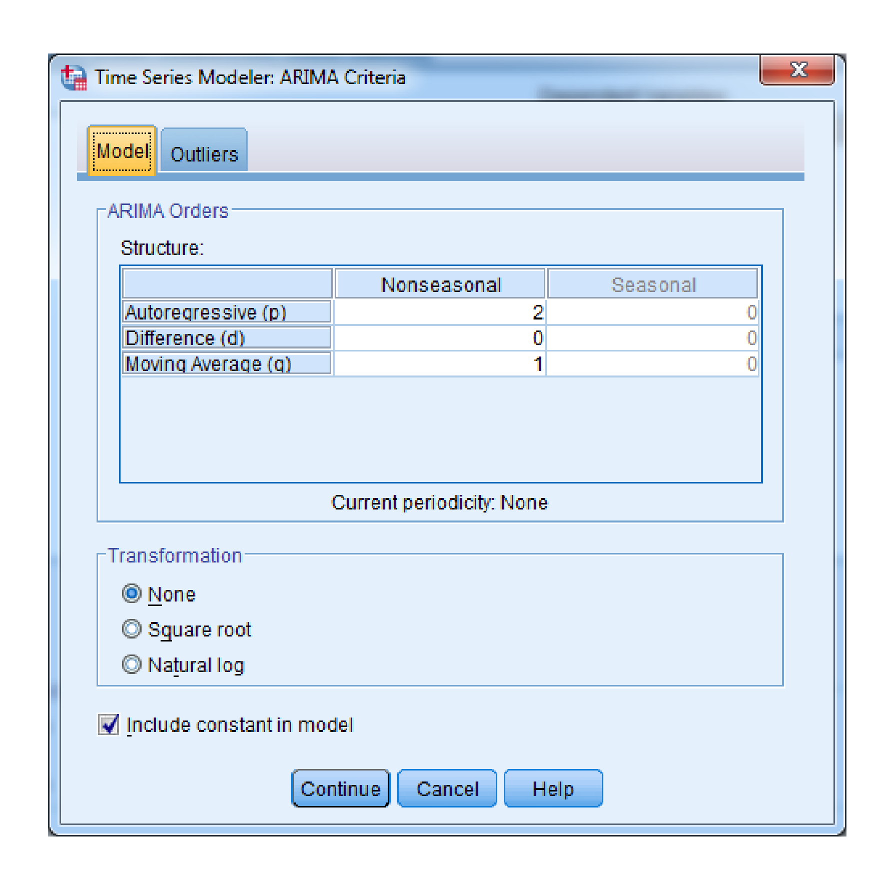



Рисунок 4.6 – Диалоговое окно «Time Series Modeler: ARIMA Criteria»

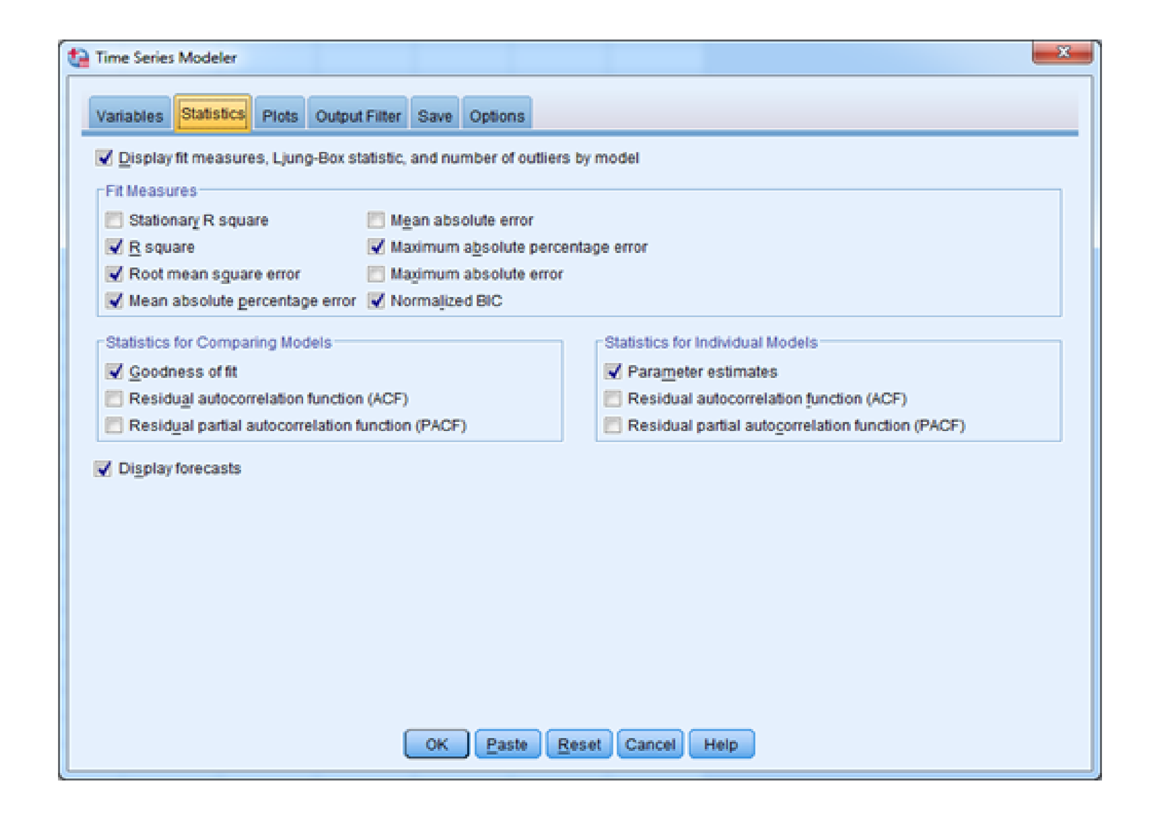



Рисунок 4.7 – Диалоговое окно «Time Series Modeler». Вкладка «Statistics»

Полученные расчетные коэффициенты моделей ARMA для первого региона представлены в табл. 4.5. Для остальных регионов расчетные значения представлены в табл. Г.1-Г.3 (Приложение Г).

Таблица 4.5

Параметры моделей ARMA для первого региона

| Модель Бокса-Дженкинса | Значения | | | Стандартная ошибка АКФ, SE | t-критерий | Значимость, P. |
|---|---|---|---|---|---|---|
| | Константа | | 7154,930 | 155,058 | 46,144 | 0,0001 |
| ARMA (2, 1) | Коэффициенты AR | Лаг 1 | 1,656 | 0,152 | 10,929 | 0,0001 |
| | | Лаг 2 | -0,943 | 0,160 | -5,893 | 0,0010 |
| | Коэффициент MA | Лаг 1 | 0,997 | 85,557 | 0,120 | 0,0090 |
| ARMA (1, 2) | Константа | | 7115,963 | 652,394 | 10,907 | 0,0001 |
| | Коэффициент AR | Лаг 1 | 0,672 | 0,466 | 1,443 | 0,0199 |
| | Коэффициенты MA | Лаг 1 | -1,045 | 8,817 | -0,119 | 0,9100 |
| | | Лаг 2 | -0,060 | 0,662 | -0,091 | 0,0310 |
| ARMA (2, 2) | Константа | | 7117,078 | 377,626 | 18,847 | 0,0001 |
| | Коэффициенты AR | Лаг 1 | 1,577 | 0,583 | 2,706 | 0,0420 |
| | | Лаг 2 | -0,793 | 0,432 | -1,837 | 0,0126 |
| | Коэффициенты MA | Лаг 1 | 0,025 | 38,746 | 0,001 | 1,0000 |
| | | Лаг 2 | 0,961 | 39,330 | 0,024 | 0,0010 |

Закономерность исследуемого процесса находит отражение в наличии определенных статистических свойств остаточной компоненты, а именно: независимости уровней, их случайности, соответствия нормальному закону распределения и равенства нулю средней ошибки.

Пусть построена некоторая модель, на основании которой вычислено значение $\hat{y}_t$. Тогда $e_t = y_t - \hat{y}_t$ – остаток в определении $y_t$, где $y_t$ – наблюдение.

Для того, чтобы модель могла считаться хорошей, необходимо доказать, что ряд остатков $e_t$ является случайным и подчиняется нормальному закону распределений. Для доказательства используется:

1) *проверка равенства нулю математического ожидания*. Выдвигается основная гипотеза

$H_0: \ \mu = 0$, то есть проверяется $\bar{e} = 0$

Для проверки этой гипотезы строится критерий или случайная величина Стьюдента:

$$t_k = r_k / SE(r_k),$$

где $SE(r_k) \approx \frac{1}{\sqrt{n}} = \sqrt{\frac{1 + 2 \cdot \sum_{i=1}^{k-1} r_i^2}{n}}$ - стандартная ошибка для $r_k$.

По таблице распределения Стьюдента при заданном $\alpha = 0{,}05$ находим $t_{кр} = t(\alpha, n)$. Если $t < t_{кр}$ – считают, что $\mu = 0$, если $t > t_{кр}$ – считают, что $\mu \neq 0$.

Оценим полученные результаты.

Сравнив полученные значения t-критерия с табличным при заданном уровне значимости α=0,05 для модели ARMA (2, 1) отвергаем гипотезу о незначимости коэффициентов модели (P< 0.05).

2) *Проверка наличия или отсутствия автокорреляции в отклонении модели*. Независимость остаточной компоненты означает отсутствие автокорреляции между остатками. Автокорреляция означает, что некоторое значение $e_t$ зависит от одного или нескольких своих предыдущих значений.

Тест Дарбина-Уотсона связан с проверкой гипотезы об отсутствии автокорреляции первого порядка, т.е. автокорреляции между соседними остаточными членами ряда. При этом статистика определяется по формуле:

$$D = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \qquad (4.2)$$

Можно показать, что величина $D \approx 2 \cdot (1 - r_1)$, где $r_1$ – коэффициент автокорреляции первого порядка (т.е. парный коэффициент корреляции между двумя последовательностями остатков $e_1, e_2, \ldots, e_{n-1}$ и $e_2, e_3, \ldots, e_n$).

Из (4.2) видно, что близость значения статистики D к нулю означает наличие высокой положительной автокорреляции (коэффициент $r_1$ близок к единице); близость значения статистики D к четырем означает наличие высокой отрицательной автокорреляции (коэффициент $r_1$ близок к минус единице). Естественно, в случае отсутствия автокорреляции значение статистики D будет близким к двум (коэффициент $r_1$ не сильно отличается от нуля).

Применение на практике критерия Дарбина-Уотсона основано на сравнении расчетного значения статистики D с пороговыми, граничными значениями $D_L$ и $D_U$. Граничные значения $D_L$ и $D_U$, зависящие от числа наблюдений n, количества объясняющих переменных в модели, уровня значимости α, находятся по таблицам с критериальными значениями.

Алгоритм выявления автокорреляции остатков на основе критерия Дарбина-Уотсона следующий.

Выдвигается гипотеза $H_0$ об отсутствии автокорреляции остатков. Пусть альтернативная гипотеза состоит в наличии в остатках положительной автокорреляции первого порядка. Тогда при сравнении расчетного значения статистики D с $D_L$ и $D_U$ возможны следующие варианты:

- если $D < D_L$, то существует положительная корреляция, гипотеза $H_0$ отвергается;
- если $D > 4 - D_L$, то наблюдается отрицательная автокорреляция, гипотеза $H_0$ не отвергается;
- если $D_U \leq D \leq 4 - D_U$, то корреляция отсутствует, гипотеза $H_0$ не отвергается;
- имеются области неопределенности: $D_L < D < D_U$, $4 - D_U \leq D \leq 4 - D_L$.

Данный критерий нельзя использовать, если среди объясняющих переменных содержатся легированные значения результативного показателя (например, он не применим к моделям авторегрессии). Таким образом, можно считать, что в случае отсутствия автокорреляции в остатках расчетное значение статистики «не слишком отличается» от 2.

Мы рассчитали интеграционную статистики Дарбина-Уотсона (IDW). Значение статистики вычисляется по формуле:

$$IDW = \frac{\sum_{t=2}^{n}(Y_t - Y_{t-1})^2}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2}$$

где $\overline{Y}$ – среднее значение ряда.

Рассчитанные значения IDW-статистики сравниваются с критическими и в случае IDW < $IDW_U$ гипотеза о стационарности ряда принимается.

Получили, например, для ряда первого региона значение интеграционной статистики Дарбина-Уотсона IDW ≈ 1,47. Сравнивая это значение с верхним критическим значением статистики $IDW_U \approx 1,6$, принимаем гипотезу о стационарности ряда остатков. Следовательно, можно сказать, что ряд остатков действительно является «белым шумом», а модель точной. Такое же поведение остатков наблюдается и для остальных регионов.

При графическом анализе учитывали, что ЧАКФ (частная автокорреляционная функция) и АКФ белого шума не должны выходить за пределы красной штриховой линии, та как остатки должны быть малыми и не должно быть закономерных корреляций.

**Этап 4. Выбор лучшей модели.**

При выборе параметров модели учитывали показатели ее точности. Они описывают величины случайных ошибок, полученных при использовании модели.

Ошибка прогноза - величина, характеризующая расхождение между фактическим и прогнозным значением показателя [57].

Абсолютная ошибка прогноза определяется по формуле:

$$\Delta_t = \hat{y}_t - y_t$$

где $\hat{y}_t$ - прогнозное значение показателя;

$y_t$ - фактическое значение.

Эта характеристика имеет ту же размерность, что и прогнозируемый показатель, и зависит от масштаба измерения уровней временного ряда.

Ошибки прогнозов, полученные ретроспективно характеризуют точность применяемой модели.

При сравнении прогнозных и действительных значений ряда рассчитали ошибки. Среди множества типов ошибок, самой подходящей является абсолютная относительная ошибка – *APE* [58]:

$$APE_k = \left| \frac{y_k - \hat{y}_k}{y_k} \right|,$$

где $y_k$ - данные, $\hat{y}_k$ - прогноз.

Среднее значение ошибки *APE* обозначается *MAPE* [59]:

$$MAPE = \frac{1}{n} \sum_{k=1}^{n} \left| \frac{y_k - \hat{y}_k}{y_k} \right| \cdot 100\%$$

Если *MAPE* < 10%, то прогноз сделан с высокой точностью, 10% < 20% - прогноз хороший, 20% < MAPE < 50% - прогноз удовлетворительный, MAPE > 50% - прогноз плохой.

MAPE подчеркивает, насколько велики ошибки прогноза в сравнении с действительными значениями ряда.

Естественно ожидать, что правильно подобранный метод даст малые ошибки прогноза.

Кроме ясности, *MAPE* обладает большей устойчивостью к выбросам по сравнению со средней квадратичной ошибкой.

Среднеквадратическая ошибка прогноза (Mean Squared Error, MSE) используется в качестве оценки метода прогнозирования. Поскольку каждое значение отклонения возводится в квадрат, то подчеркиваются большие ошибки прогноза:

$$MSE = \frac{1}{n} \sum_{k=1}^{n} (y_k - \tilde{y}_k)^2$$

Следовательно, прогнозы на основе *MAPE* будут более надежны. Модель ряда считается хорошей, если ошибка прогноза *MAPE* мала и остатки ряда: $остатки = данные - прогноз$ имеют вид белого шума, т.е. их

дисперсия постоянна и они некоррелированы [60 – 63].

Качество моделей сравнивалось по информационному критерию Шварца ($SBC_G$ ).

Байесовский критерий Шварца (Normalized BIC) для моделей регрессии можно представить в виде:

$$SBC_G = \ln(\tilde{G}^2) + (\frac{\ln n}{n}) \cdot p,$$

где n – объём выборочной совокупности;

p – число параметров в модели .

$\tilde{G}^2$ ($MSE^2$) – оценка максимального правдоподобия дисперсии остатков $e_t$ модели регрессии.

Выбирается та модель, для которой значение $SBC_G$ является наименьшим.

В пакете SPSS для получения показателей качества моделей ARMA (1, 2), (2, 1) и (2, 2) в диалоговом окне «Time Series Modeler». Вкладка «Statistics» (рис. 4.7) выбрали основные критерии их оценки, а именно «R squre», «Root mean square error», «Mean absolute percentage error», «Maximum absolute percentage error», «Normalized BIC». Результаты для первого региона приведены в табл. 4.6.

Таблица 4.6

Основные критерии оценки качества моделей ARMA

| Модель | Критерии оценки качества | | | | |
|---|---|---|---|---|---|
| | $R^2$ | MSE | MAPE | MaxAPE | $SBC_G$ |
| ARMA (1, 2) | 0,804 | 470,577 | 4,033 | 11,343 | 13,229 |
| ARMA (2, 2) | 0,835 | 472,470 | 3,841 | 8,655 | 13,467 |
| ARMA (2, 1) | 0,836 | 430,458 | 4,297 | 8,646 | 13,051 |

Коэффициент детерминации $R^2$ во всех трех моделях выше 0,8, т.е. модели работают хорошо (имеют высокую значимость). Анализируя полученные значения MSE, MAPE, MaxAPE и $SBC_G$ для четырех регионов Украины можно сделать вывод, что модель ARMA (2,1) является лучшей.

Таким образом, для прогнозирования заболеваемости сельского населения НПР для 1-4 регионов предлагаем использовать следующие 4 модели временного ряда:

- для первого региона

$$\hat{y}_t = 7154.930 + 1.656 \cdot y_{t-1} - 0.943 \cdot y_{t-2} + 0.997 \cdot \varepsilon_{t-1};$$

- для второго региона

$$\hat{y}_t = 7713.684 + 1.677 \cdot y_{t-1} - 0.952 \cdot y_{t-2} + 0.993 \cdot \varepsilon_{t-1};$$

- для третьего региона

$$\hat{y}_t = 6040.129 + 0.745 \cdot y_{t-1} - 0.116 \cdot y_{t-2} - 0.993 \cdot \varepsilon_{t-1};$$

- для четвертого региона

$$\hat{y}_t = 3154.232 + 0.780 \cdot y_{t-1} - 0.084 \cdot y_{t-2} - 0.999 \cdot \varepsilon_{t-1}.$$

Полученные модели можно признать удачными, все коэффициенты в модели являются статистически значимыми ($P < 0.05$), в модели присутствует автокорреляция, остатки моделей стационарны и имеют нормальное распределение.

Оценку АКФ и ЧАКФ остатков, получили в SPSS (рис. 4.8) путем установки в диалоговом окне «Time Series Modeler». Вкладка «Plot»опций «Residual autocorrelation function (ACF)» и «Residual partial autocorrelation function (PACF)».

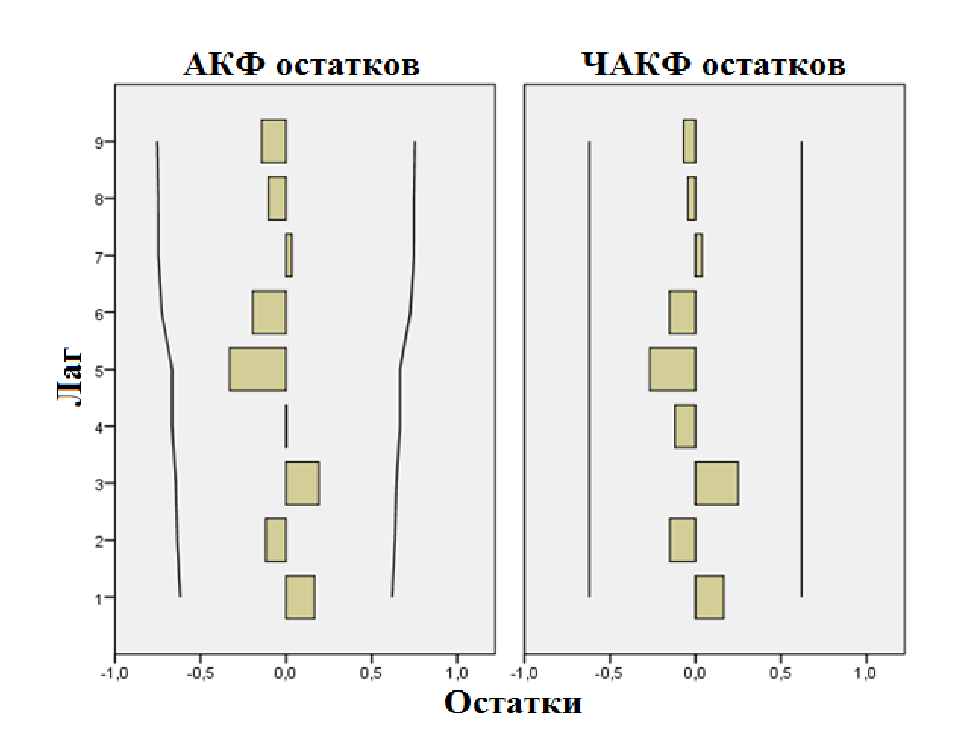



Рисунок 4.8 – АКФ и ЧАКФ остатков для 1-го региона

Из рисунка следует, что ряд остатков является «белым шумом», а построенная модель прогноза для первого региона является точной.

Модели прогноза для остальных регионов также являются точными рис. Г1 – Г3 (Приложение Г).

Анализ и верификация прогностических моделей показали, что точность прогнозирования достаточно высокая, полученные результаты приемлемы для принятия управленческих решений в процедуре выбора и формирования лечебно- профилактических мероприятий

**Этап 6.** Прогнозирование с помощью построенной модели.

Было проведено прогнозирование заболеваемости непсихотическими психическими расстройствами седьского населения Украины с использованием синтезированных моделей.

В пакете SPSS для получения показателей прогнозирования по заболеваемости непсихотическими психическими расстройствами по каждому региону в диалоговом окне «Time Series Modeler». Вкладка «Statistics» (рис. 4.7) установили флажок напротив опции «Display forecasts».

Результаты прогнозирования по заболеваемости непсихотическими психическими расстройствами представлены в табл. 4.7.

Для построения графиков прогноза заболеваемости сельского населения непсихотическими психическими расстройствами по всем 4-м регионам Украины в диалоговом окне «Time Series Modeler» (рис. 4.9) перешли на вкладку «Plot» и установили отметку в соответствующем поле «Plots for individual Models». После этого нажали кнопку «ОК».

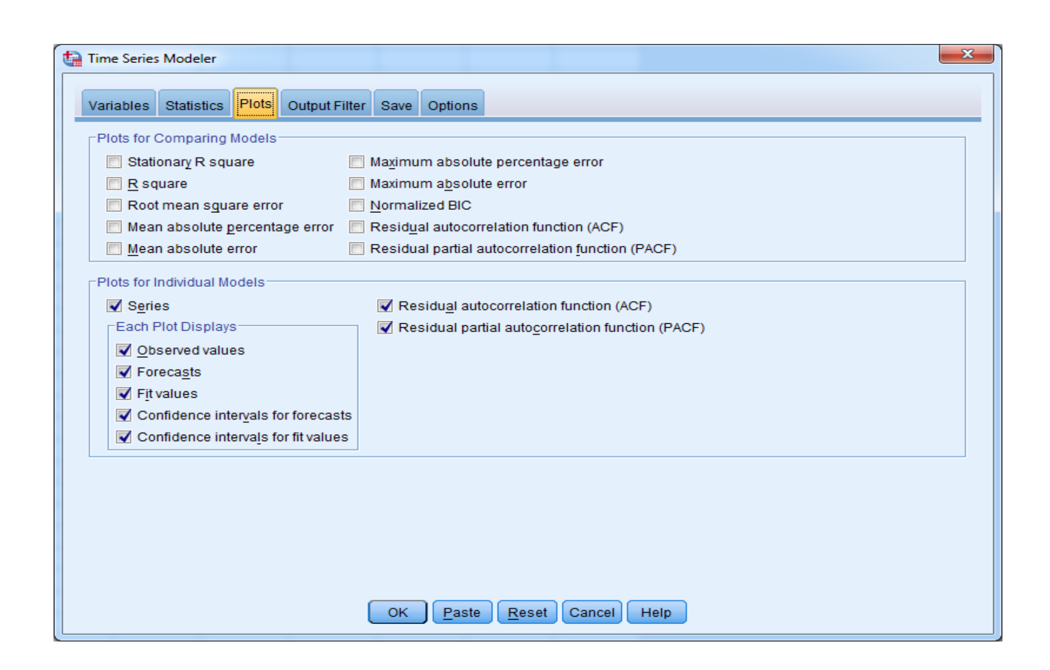



Рисунок 4.9 – Диалоговое окно «Time Series Modeler». Вкладка «Plot»

Из полученных графиков (рис. 4.10 – 4.13) видно, что прогнозные кривые практически повторяют фрагменты кривой исходного ряда, причём все наблюдаемые значения ряда попадают в доверительный интервал.

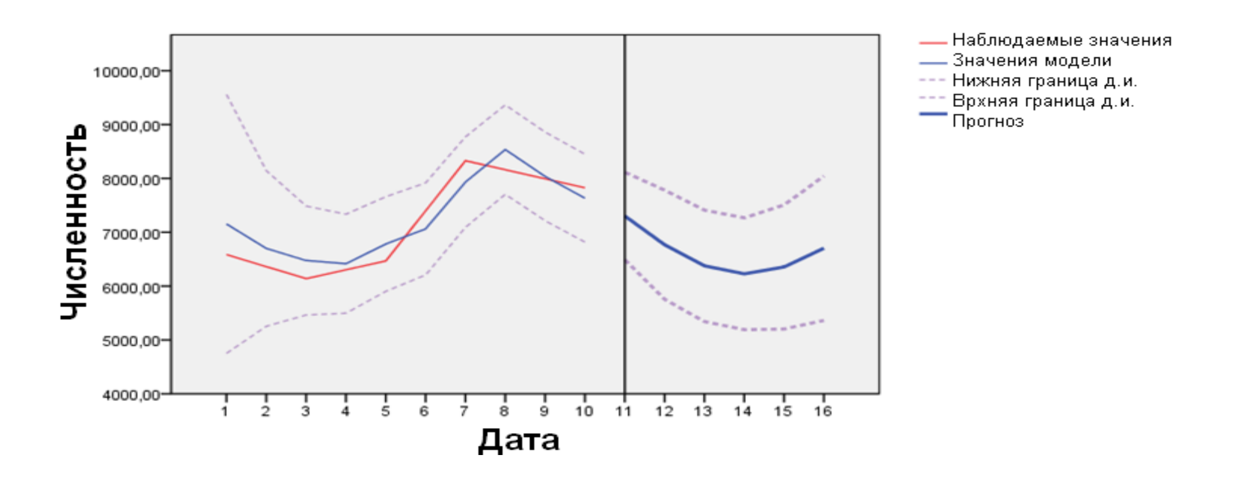



Рисунок 4.10 – Прогноз заболеваемости сельского населения Украины непсихотическими психическими расстройствами в первом регионе

Таблица 4.7

Результаты прогнозирования по заболеваемости непсихотическими психическими расстройствами для четырех регионов Украины

| Регион | Год | 2013 | | 2014 | | 2015 | |
|---|---|---|---|---|---|---|---|
| 1 | Верхний доверительный интервал | 8115,71 | 7782,70 | 7411,10 | 7267,90 | 7507,82 | 8044,50 |
| | Прогноз | 7303,62 | 6769,51 | 6376,50 | 6229,16 | 6355,62 | 6703,95 |
| | Нижний доверительный интервал | 6491,53 | 5756,33 | 5341,91 | 5190,42 | 5203,43 | 5363,41 |
| 2 | Верхний доверительный интервал | 8973,37 | 8599,03 | 8108,83 | 7793,88 | 7864,12 | 8301,01 |
| | Прогноз | 8159,07 | 7572,57 | 7052,92 | 6740,17 | 6710,67 | 6959,06 |
| | Нижний доверительный интервал | 7344,78 | 6546,12 | 5997,01 | 5686,45 | 5557,22 | 5617,12 |
| 3 | Верхний доверительный интервал | 7739,20 | 8566,31 | 8769,49 | 8796,40 | 8785,53 | 8772,56 |
| | Прогноз | 6586,51 | 6341,42 | 6201,01 | 6124,92 | 6084,58 | 6063,38 |
| | Нижний доверительный интервал | 6433,83 | 4916,52 | 4632,53 | 4653,45 | 4683,63 | 4854,20 |
| 4 | Верхний доверительный интервал | 3615,77 | 3457,89 | 3352,32 | 3283,24 | 3238,22 | 3208,91 |
| | Прогноз | 4333,66 | 4865,15 | 5013,87 | 5045,54 | 5041,86 | 5029,81 |
| | Нижний доверительный интервал | 2897,89 | 2050,63 | 2690,77 | 1920,94 | 2434,58 | 2388,01 |

Таким образом, на основании полученных результатов следует отметить, что заболеваемость непсихотическим психическими расстройствами у сельского населения на протяжении 2011 - 2012 гг. носила отрицательный характер, то есть наметился спад уровня заболеваемости по всем регионам Украины.

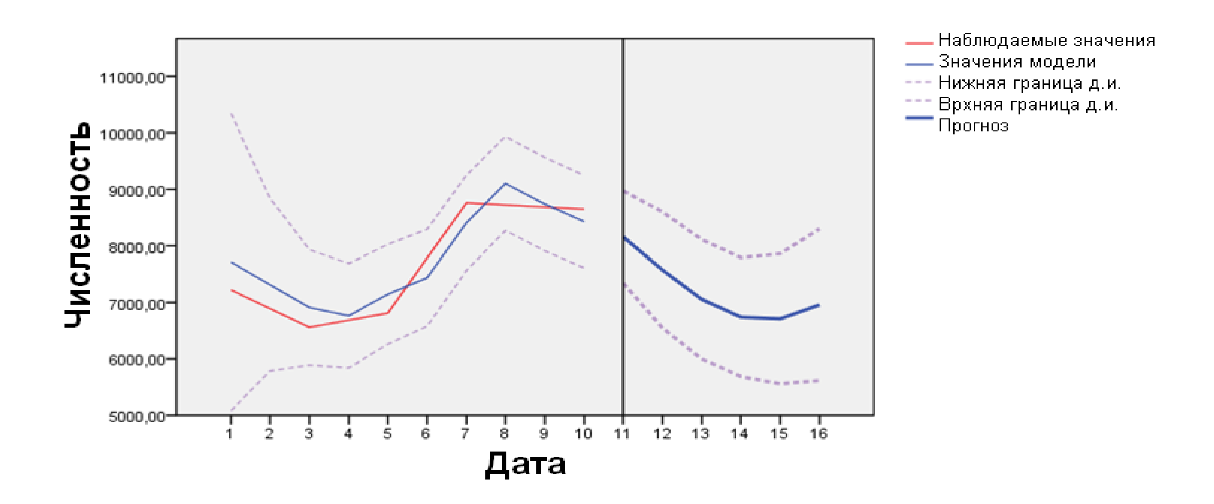



Рисунок 4.11 – Прогноз заболеваемости сельского населения Украины непсихотическими психическими расстройствами во втором регионе на 2015 год

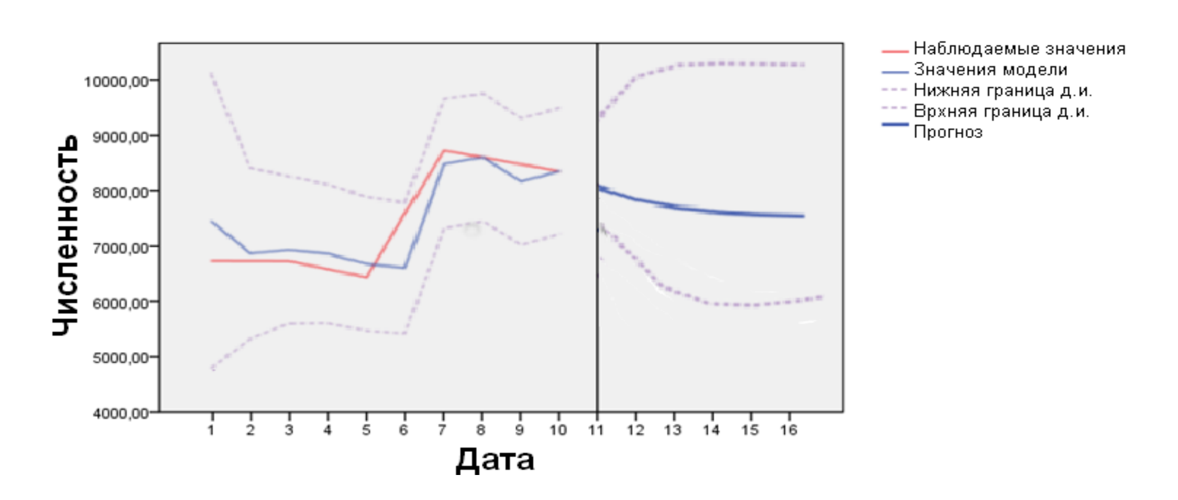



Рисунок 4.12 – Прогноз заболеваемости сельского населения Украины непсихотическими психическими расстройствами в третьем регионе на 2015 год

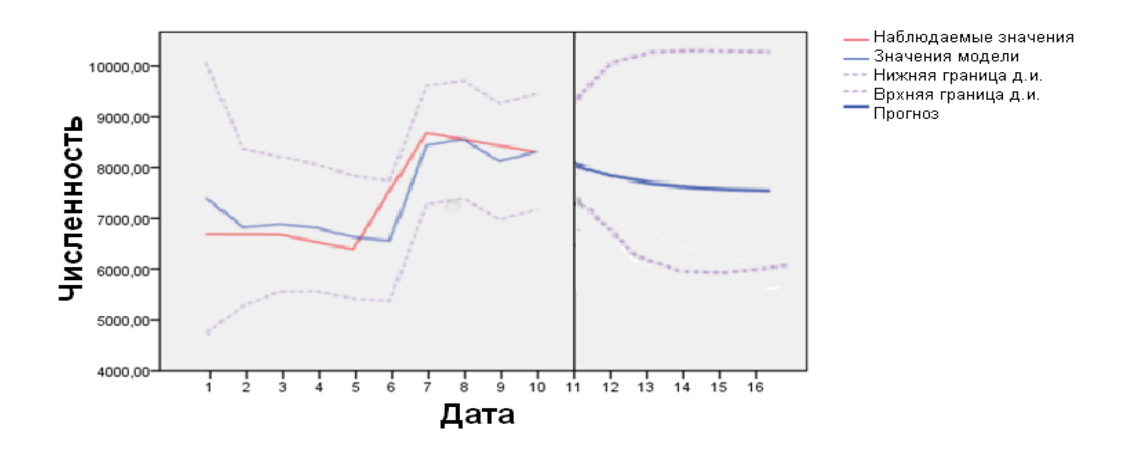



Рисунок 4.13. – Прогноз заболеваемости сельского населения Украины непсихотическими психическими расстройствами в четвертом регионе на 2015 год

Для оптимизации медицинского обеспечения лиц, страдающих непсихотическими психическими расстройствами, необходимо пользоваться прогностическими моделями заболеваемости, на основе которых должно проводиться планирование ресурсов и объемов медицинской помощи, а также дальнейшее усовершенствование психиатрической службы, направленное на повышение качества, территориальной и экономической доступности психиатрической помощи сельскому населению Украины.

## ПРИЛОЖЕНИЕ А

Таблица А. 1

Кодировочная таблица к задаче «Способ диагностики хронической сердечной недостаточности у подростков на ранней стадии развития»

| № | Показатель | Описание | Код показателя |
|---|---|---|---|
| 1 | diagnoz | Диагноз | 1- миокардиофиброз<br>2- дизпластическая кардимиоопатия<br>3- гипертрофическая кардимиопатия<br>4- дилятационная кардимиопатия<br>5- аритмия<br>6- синдром перенапряжения миокарда<br>7- коарктация аорты<br>8- стеноз аорты<br>9- стеноз легочной артерии<br>10- дефект межжелудочковой перегородки<br>11- дефект межпредсердной перегородки<br>12- сочетанные пороки<br>13- тетрада Фалло<br>14- здоров |
| 2 | Diagnoz1 | Диагноз | 1- миокардиофиброз<br>2- дизпластическая кардиомиопатия<br>3- пороки с перегрузкой левых отделов сердца<br>4- пороки с перегрузкой правых отделов сердца<br>5- порки по смешанному типу<br>6- синдром перенапряжения миокарда<br>7- условно здоровые |
| 3 | sn | Сердечная недостаточность | 1- без сердечной недостаточности<br>2- 1-я стадия сердечной недостаточности<br>3- контрольная группа |
| 4 | pol | Пол | 1- М<br>2- Ж |
| 5 | im_srb | ИММУНОЛОГИЯ С-реактивный белок | 0 – отрицательный<br>1- Положительный (определяется) |
| 6 | im_asl_o | ИММУНОЛОГИЯ антистрептолизин-О; абс.ед. | 0 – норма ( 0 – 200)<br>1 – более 200 |

| Сердечная недостаточность | ЦИТОКИНЫ IL 1, пг/мл | ЦИТОКИНЫ IL 6, пг/мл | ЦИТОКИНЫ ФНО а, пг/мл | СД-95, % клеток апоптотичных | цАМФ, циклический уденозинмонофосфат, нмоль/мл | РЕНИН, нг/мл/час | КАРБОНИЛИРОВАННЫЕ БЕЛКИ, у.е. | НОРАДРЕНАЛИН, нмоль/сут | ЭХОКАРДИОГРАФИЯ ПОСЛЕ НАГРУЗКИ ФВ, % | ИКДО, индекс конечного диастолического объема лев.жел., у.е. |
|---|---|---|---|---|---|---|---|---|---|---|
| Итого ЦИТОКИНЫ IL 1, пг/мл | 271,605 | -,053 | 81,314 | 13,230 | 51,001 | -,685 | 1,087 | -99,283 | 152,184 | 23,081 |
| ЦИТОКИНЫ IL 6, пг/мл | -,053 | 22,138 | 6,854 | 4,535 | -1,276 | ,282 | ,047 | 12,719 | 43,192 | 1,131 |
| ЦИТОКИНЫ ФНО а, пг/мл | 81,314 | 6,854 | 164,842 | 15,506 | 36,783 | ,678 | ,035 | 35,963 | 177,152 | 4,071 |
| СД-95, % клеток апоптотичных | 13,230 | 4,535 | 15,506 | 16,485 | 2,704 | -,243 | ,165 | 25,412 | 60,057 | ,979 |
| цАМФ, циклический уденозинмонофосфат, нмоль/мл | 51,001 | -1,276 | 36,783 | 2,704 | 114,967 | -1,749 | ,218 | 43,346 | 15,238 | 3,151 |
| РЕНИН, нг/мл/час | -,685 | ,282 | ,678 | -,243 | -1,749 | ,365 | ,031 | 2,207 | 2,047 | 1,020 |
| КАРБОНИЛИРОВАННЫЕ БЕЛКИ, у.е. | 1,087 | ,047 | ,035 | ,165 | ,218 | ,031 | ,215 | -2,175 | 2,051 | ,523 |
| НОРАДРЕНАЛИН, нмоль/сут | -99,283 | 12,719 | 35,963 | 25,412 | 43,346 | 2,207 | -2,175 | 1695,138 | 121,159 | 52,097 |
| ЭХОКАРДИОГРАФИЯ ПОСЛЕ НАГРУЗКИ ФВ, % | 152,184 | 43,192 | 177,152 | 60,057 | 15,238 | 2,047 | 2,051 | 121,159 | 932,818 | 39,221 |
| ИКДО, индекс конечного диастолического объема лев. жел., у.е. | 23,081 | 1,131 | 4,071 | ,979 | 3,151 | 1,020 | ,523 | 52,097 | 39,221 | 165,189 |

Рисунок А.1 – Общая ковариационная матрица *H*

| Сердечная недостаточность | | ЦИТОКИНЫ IL 1, пг/мл | ЦИТОКИНЫ IL 6, пг/мл | ЦИТОКИНЫ ФНО а, пг/мл | СД-95, % клеток апоптотичных | цАМФ, циклический уденозинмонофосфат, нмоль/мл | РЕНИН, нг/мл/час | КАРБОНИЛИРОВАННЫЕ БЕЛКИ, у.е. | НОРАДРЕНАЛИН, нмоль/сут | ЭХОКАРДИОГРАФИЯ ПОСЛЕ НАГРУЗКИ ФВ, % | ИКДО, индекс конечного диастолического объема лев.жел., у.е. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Итого | ЦИТОКИНЫ IL 1, пг/мл | 161,866 | 2,021 | 50,597 | -1,711 | 58,164 | -1,150 | ,079 | 14,806 | 36,141 | 3,414 |
| | ЦИТОКИНЫ IL 6, пг/мл | 2,021 | 20,126 | ,493 | 2,415 | -1,989 | ,184 | ,030 | -3,749 | 4,937 | -,677 |
| | ЦИТОКИНЫ ФНО а, пг/мл | 50,597 | ,493 | 136,156 | 3,824 | 37,282 | ,215 | -,362 | 23,582 | 18,764 | -8,289 |
| | СД-95, % клеток апоптотичных | -1,711 | 2,415 | 3,824 | 12,013 | 2,980 | -,427 | -,008 | 25,353 | ,550 | -4,051 |
| | цАМФ, циклический уденозинмонофосфат, нмоль/мл | 58,164 | -1,989 | 37,282 | 2,980 | 115,919 | -1,774 | ,271 | 33,556 | 11,735 | 3,779 |
| | РЕНИН, нг/мл/час | -1,150 | ,184 | ,215 | -,427 | -1,774 | ,363 | ,025 | 1,991 | -,403 | ,846 |
| | КАРБОНИЛИРОВАННЫЕ БЕЛКИ, у.е. | ,079 | ,030 | -,362 | -,008 | ,271 | ,025 | ,208 | -1,390 | ,366 | ,315 |
| | НОРАДРЕНАЛИН, нмоль/сут | 14,806 | -3,749 | 23,582 | 25,353 | 33,556 | 1,991 | -1,390 | 1505,989 | -20,465 | 59,039 |
| | ЭХОКАРДИОГРАФИЯ ПОСЛЕ НАГРУЗКИ ФВ, % | 36,141 | 4,937 | 18,764 | ,550 | 11,735 | -,403 | ,366 | -20,465 | 78,493 | -20,778 |
| | ИКДО, индекс конечного диастолического объема лев. жел., у.е. | 3,414 | -,677 | -8,289 | -4,051 | 3,779 | ,846 | ,315 | 59,039 | -20,778 | 161,667 |

Рисунок А.2 – Внутригрупповая ковариационная матрица *A*

| Сердечная недостаточность | | ЦИТОКИНЫ IL 1, пг/мл | ЦИТОКИНЫ IL 6, пг/мл | ЦИТОКИНЫ ФНО а, пг/мл | СД-95, % клеток апоптотичных | цАМФ, циклический уденозинмонофосфат, нмоль/мл | РЕНИН, нг/мл/час | КАРБОНИЛИРОВАННЫЕ БЕЛКИ, у.е. | НОРАДРЕНАЛИН, нмоль/сут | ЭХОКАРДИОГРАФИЯ ПОСЛЕ НАГРУЗКИ ФВ, % | ИКДО, индекс конечного диастолического объема лев.жел., у.е. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Итого | ЦИТОКИНЫ IL 1, пг/мл | 1,000 | ,035 | ,341 | -,039 | ,425 | -,150 | ,014 | ,030 | ,321 | ,021 |
| | ЦИТОКИНЫ IL 6, пг/мл | ,035 | 1,000 | ,009 | ,155 | -,041 | ,068 | ,015 | -,022 | ,124 | -,012 |
| | ЦИТОКИНЫ ФНО а, пг/мл | ,341 | ,009 | 1,000 | ,095 | ,297 | ,031 | -,068 | ,052 | ,182 | -,056 |
| | СД-95, % клеток апоптотичных | -,039 | ,155 | ,095 | 1,000 | ,080 | -,205 | -,005 | ,188 | ,018 | -,092 |
| | цАМФ, циклический уденозинмонофосфат, нмоль/мл | ,425 | -,041 | ,297 | ,080 | 1,000 | -,274 | ,055 | ,080 | ,123 | ,028 |
| | РЕНИН, нг/мл/час | -,150 | ,068 | ,031 | -,205 | -,274 | 1,000 | ,092 | ,085 | -,075 | ,110 |
| | КАРБОНИЛИРОВАННЫЕ БЕЛКИ, у.е. | ,014 | ,015 | -,068 | -,005 | ,055 | ,092 | 1,000 | -,079 | ,090 | ,054 |
| | НОРАДРЕНАЛИН, нмоль/сут | ,030 | -,022 | ,052 | ,188 | ,080 | ,085 | -,079 | 1,000 | -,060 | ,120 |
| | ЭХОКАРДИОГРАФИЯ ПОСЛЕ НАГРУЗКИ ФВ, % | ,321 | ,124 | ,182 | ,018 | ,123 | -,075 | ,090 | -,060 | 1,000 | -,184 |
| | ИКДО, индекс конечного диастолического объема лев. жел., у.е. | ,021 | -,012 | -,056 | -,092 | ,028 | ,110 | ,054 | ,120 | -,184 | 1,000 |

Рисунок А.3 – Объединенная внутригрупповая матрица $Q$

| Сердечная недостаточность | | ЦИТОКИНЫ IL 1, пг/мл | ЦИТОКИНЫ IL 6, пг/мл | ЦИТОКИНЫ ФНО а, пг/мл | СД-95, % клеток апоптотичных | цАМФ, циклический уденозинмонофосфат, нмоль/мл | РЕНИН, нг/мл/час | КАРБОНИЛИРОВАННЫЕ БЕЛКИ, у.е. | НОРАДРЕНАЛИН, нмоль/сут | ЭХОКАРДИОГРАФИЯ ПОСЛЕ НАГРУЗКИ ФВ, % | ИКДО, индекс конечного диастолического объема лев.жел., у.е. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Без сердечной недостаточности | ЦИТОКИНЫ IL 1, пг/мл | 872,800 | 15,798 | 242,098 | -14,008 | 281,766 | -6,358 | ,454 | 80,305 | 186,777 | 63,549 |
| | ЦИТОКИНЫ IL 6, пг/мл | 15,798 | 17,066 | 15,316 | 3,915 | 5,830 | -,524 | -,073 | 7,975 | 13,875 | -10,408 |
| | ЦИТОКИНЫ ФНО а, пг/мл | 242,098 | 15,316 | 364,255 | 10,996 | 130,568 | -3,203 | -2,467 | -4,229 | 98,819 | 30,286 |
| | СД-95, % клеток апоптотичных | -14,008 | 3,915 | 10,996 | 19,169 | 10,250 | -1,139 | -,039 | 46,894 | -4,865 | -11,062 |
| | цАМФ, циклический уденозинмонофосфат, нмоль/мл | 281,766 | 5,830 | 130,568 | 10,250 | 198,114 | -3,910 | -,324 | 26,904 | 54,290 | 50,200 |
| | РЕНИН, нг/мл/час | -6,358 | -,524 | -3,203 | -1,139 | -3,910 | ,469 | -,040 | -2,311 | -,260 | ,878 |
| | КАРБОНИЛИРОВАННЫЕ БЕЛКИ, у.е. | ,454 | -,073 | -2,467 | -,039 | -,324 | -,040 | ,301 | -1,351 | ,975 | ,457 |
| | НОРАДРЕНАЛИН, нмоль/сут | 80,305 | 7,975 | -4,229 | 46,894 | 26,904 | -2,311 | -1,351 | 562,787 | 37,902 | -65,347 |
| | ЭХОКАРДИОГРАФИЯ ПОСЛЕ НАГРУЗКИ ФВ, % | 186,777 | 13,875 | 98,819 | -4,865 | 54,290 | -,260 | ,975 | 37,902 | 157,019 | -22,583 |
| | ИКДО, индекс конечного диастолического объема лев. жел., у.е. | 63,549 | -10,408 | 30,286 | -11,062 | 50,200 | ,878 | ,457 | -65,347 | -22,583 | 149,372 |
| 1 стадия сердечной недостаточности | ЦИТОКИНЫ IL 1, пг/мл | 35,004 | -2,504 | 25,026 | ,832 | 24,794 | -,184 | -,017 | 4,525 | 11,004 | -21,981 |
| | ЦИТОКИНЫ IL 6, пг/мл | -2,504 | 39,515 | -4,956 | 4,218 | -7,048 | ,649 | ,051 | -11,015 | 7,628 | -,256 |
| | ЦИТОКИНЫ ФНО а, пг/мл | 25,026 | -4,956 | 181,164 | 4,505 | 38,202 | 1,910 | ,048 | 62,922 | 5,215 | -32,262 |
| | СД-95, % клеток апоптотичных | ,832 | 4,218 | 4,505 | 10,476 | 7,840 | -,464 | -,074 | 34,513 | 1,615 | -6,067 |
| | цАМФ, циклический уденозинмонофосфат, нмоль/мл | 24,794 | -7,048 | 38,202 | 7,840 | 138,709 | -2,457 | ,783 | 70,012 | 4,512 | -24,597 |
| | РЕНИН, нг/мл/час | -,184 | ,649 | 1,910 | -,464 | -2,457 | ,576 | ,071 | 4,930 | -,858 | 1,814 |
| | КАРБОНИЛИРОВАННЫЕ БЕЛКИ, у.е. | -,017 | ,051 | ,048 | -,074 | ,783 | ,071 | ,213 | -3,696 | ,781 | -,067 |
| | НОРАДРЕНАЛИН, нмоль/сут | 4,525 | -11,015 | 62,922 | 34,513 | 70,012 | 4,930 | -3,696 | 2688,478 | -128,801 | 158,038 |
| | ЭХОКАРДИОГРАФИЯ ПОСЛЕ НАГРУЗКИ ФВ, % | 11,004 | 7,628 | 5,215 | 1,615 | 4,512 | -,858 | ,781 | -128,801 | 78,973 | -26,215 |
| | ИКДО, индекс конечного диастолического объема лев. жел., у.е. | -21,981 | -,256 | -32,262 | -6,067 | -24,597 | 1,814 | -,067 | 158,038 | -26,215 | 240,147 |
| Контрольная группа | ЦИТОКИНЫ IL 1, пг/мл | 4,818 | ,981 | -,046 | ,668 | 2,437 | -,034 | ,024 | -1,027 | 1,174 | 4,349 |
| | ЦИТОКИНЫ IL 6, пг/мл | ,981 | 2,529 | -,062 | ,075 | -,165 | ,011 | ,052 | -1,322 | -1,193 | 2,748 |
| | ЦИТОКИНЫ ФНО а, пг/мл | -,046 | -,062 | 2,655 | ,339 | -,359 | -,081 | ,069 | -3,611 | ,365 | -,238 |
| | СД-95, % клеток апоптотичных | ,668 | ,075 | ,339 | 10,683 | -4,597 | -,111 | ,068 | 7,986 | 1,651 | ,665 |
| | цАМФ, циклический уденозинмонофосфат, нмоль/мл | 2,437 | -,165 | -,359 | -4,597 | 61,440 | -,270 | ,009 | ,824 | 1,975 | 13,008 |
| | РЕНИН, нг/мл/час | -,034 | ,011 | -,081 | -,111 | -,270 | ,114 | ,006 | ,835 | -,018 | -,105 |
| | КАРБОНИЛИРОВАННЫЕ БЕЛКИ, у.е. | ,024 | ,052 | ,069 | ,068 | ,009 | ,006 | ,167 | ,830 | -,277 | ,630 |
| | НОРАДРЕНАЛИН, нмоль/сут | -1,027 | -1,322 | -3,611 | 7,986 | ,824 | ,835 | ,830 | 730,897 | 61,595 | 12,039 |
| | ЭХОКАРДИОГРАФИЯ ПОСЛЕ НАГРУЗКИ ФВ, % | 1,174 | -1,193 | ,365 | 1,651 | 1,975 | -,018 | -,277 | 61,595 | 47,093 | -14,794 |
| | ИКДО, индекс конечного диастолического объема лев. жел., у.е. | 4,349 | 2,748 | -,238 | ,665 | 13,008 | -,105 | ,630 | 12,039 | -14,794 | 90,408 |

Рисунок А.4 – Групповая ковариационная матрица *R*

Таблица А. 2

Фактическая и прогнозируемые группы, значения дискриминантных функций

| Номер наблюдения | | | Наивероятнейшая группа | | | | | Вторая вероятнейшая группа | | | Значения дискриминантных функций | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | P(D>d \| G=g) | | | | | | | | |
| | | Фак-тическая группа | Предсказанная груп-па | p | ст.св | P(G=g \| D=d) | Квадрат расстоя-ния Махалонобиса до центра | Группа | P(G=g \| D=d) | Квадрат расстоя-ния Махалонобиса до центра | Функ-ция 1 | Функ-ция 2 |
| **1** | | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| Исходные | 1 | 1 | 1 | ,634 | 2 | ,987 | ,911 | 2 | ,013 | 11,351 | 3,040 | 2,534 |
| | 2 | 1 | 1 | ,015 | 2 | ,851 | 8,440 | 2 | ,115 | 14,206 | -,123 | ,502 |
| | 3 | 1 | 1 | ,437 | 2 | ,842 | 1,657 | 2 | ,158 | 6,764 | 1,704 | 1,149 |
| | 4 | 1 | 1 | ,390 | 2 | ,969 | 1,883 | 2 | ,031 | 10,500 | 1,088 | 1,446 |
| | 5 | 1 | 1 | ,290 | 2 | ,883 | 2,475 | 2 | ,117 | 8,280 | 1,229 | 1,063 |
| | 6 | 1 | 2** | ,260 | 2 | ,794 | 2,692 | 1 | ,206 | 3,631 | 3,188 | ,811 |
| | 7 | 1 | 1 | ,144 | 2 | ,680 | 3,875 | 2 | ,320 | 7,143 | 1,167 | ,640 |
| | 8 | 1 | 1 | ,223 | 2 | ,612 | 3,000 | 2 | ,388 | 5,666 | 3,625 | 1,555 |
| | 9 | 1 | 1 | ,004 | 2 | 1,000 | 11,270 | 2 | ,000 | 43,840 | 2,355 | 5,721 |
| | 10 | 1 | 1 | ,688 | 2 | ,996 | ,748 | 2 | ,004 | 13,676 | 2,829 | 2,837 |
| | 11 | 1 | 1 | ,435 | 2 | ,706 | 1,664 | 2 | ,294 | 5,177 | 2,962 | 1,416 |
| | 12 | 1 | 1 | ,014 | 2 | 1,000 | 8,604 | 2 | ,000 | 34,716 | 3,244 | 5,074 |
| | 13 | 1 | 1 | ,483 | 2 | ,861 | 1,455 | 2 | ,139 | 6,863 | 3,155 | 1,792 |
| | 14 | 1 | 2** | ,271 | 2 | ,975 | 2,612 | 1 | ,025 | 8,162 | 1,830 | -,470 |
| | 15 | 1 | 1 | ,074 | 2 | ,876 | 5,202 | 2 | ,124 | 10,867 | ,504 | ,744 |
| | 16 | 1 | 1 | ,452 | 2 | ,959 | 1,589 | 2 | ,041 | 9,643 | 3,356 | 2,289 |
| | 17 | 1 | 1 | ,000 | 2 | 1,000 | 27,464 | 3 | ,000 | 50,053 | -1,342 | 6,327 |
| | 18 | 1 | 2** | ,260 | 2 | ,885 | 2,692 | 1 | ,115 | 5,019 | 2,070 | ,134 |
| | 19 | 1 | 1 | ,070 | 2 | 1,000 | 5,327 | 2 | ,000 | 31,201 | ,309 | 3,831 |
| | 20 | 1 | 1 | ,255 | 2 | ,947 | 2,730 | 2 | ,053 | 10,245 | 3,751 | 2,367 |
| | 21 | 1 | 1 | ,061 | 2 | 1,000 | 5,584 | 2 | ,000 | 26,829 | 3,373 | 4,364 |
| | 22 | 1 | 1 | ,176 | 2 | 1,000 | 3,475 | 2 | ,000 | 25,732 | 2,547 | 4,184 |
| | 23 | 1 | 1 | ,924 | 2 | ,998 | ,157 | 2 | ,002 | 14,148 | 1,748 | 2,560 |

Продолжение таблицы А. 2

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 24 | 1 | 1 | ,112 | 2 | 1,000 | 4,371 | 2 | ,000 | 27,027 | 2,792 | 4,347 |
| 25 | 1 | 1 | ,217 | 2 | ,570 | 3,054 | 2 | ,430 | 5,379 | 1,642 | ,687 |
| 26 | 1 | 1 | ,584 | 2 | ,990 | 1,075 | 2 | ,010 | 11,957 | 3,104 | 2,629 |
| 27 | 1 | 1 | ,414 | 2 | 1,000 | 1,766 | 2 | ,000 | 22,802 | 1,517 | 3,569 |
| 28 | 2 | 2 | ,697 | 2 | ,986 | ,721 | 1 | ,014 | 7,453 | 2,753 | -,276 |
| 29 | 2 | 2 | ,236 | 2 | ,990 | 2,887 | 1 | ,010 | 10,362 | 4,621 | ,375 |
| 30 | 2 | 2 | ,669 | 2 | ,977 | ,803 | 1 | ,023 | 6,513 | 3,054 | ,008 |
| 31 | 2 | 2 | ,883 | 2 | ,994 | ,248 | 1 | ,006 | 8,860 | 3,366 | -,319 |
| 32 | 2 | 2 | ,622 | 2 | 1,000 | ,950 | 1 | ,000 | 17,205 | 2,964 | -1,683 |
| 33 | 2 | 2 | ,005 | 2 | ,956 | 10,423 | 1 | ,034 | 15,317 | ,188 | -1,042 |
| 34 | 2 | 2 | ,499 | 2 | 1,000 | 1,389 | 1 | ,000 | 19,511 | 3,053 | -1,939 |
| 35 | 2 | 2 | ,644 | 2 | ,975 | ,880 | 1 | ,025 | 6,441 | 3,241 | ,108 |
| 36 | 2 | 2 | ,057 | 2 | ,999 | 5,723 | 1 | ,001 | 17,879 | 1,211 | -1,760 |
| 37 | 2 | 2 | ,868 | 2 | 1,000 | ,284 | 1 | ,000 | 15,766 | 3,504 | -1,339 |
| 38 | 2 | 2 | ,264 | 2 | 1,000 | 2,661 | 1 | ,000 | 23,121 | 4,858 | -1,564 |
| 39 | 2 | 2 | ,818 | 2 | 1,000 | ,401 | 1 | ,000 | 15,684 | 3,227 | -1,422 |
| 40 | 2 | 2 | ,801 | 2 | ,990 | ,444 | 1 | ,010 | 7,839 | 3,031 | -,266 |
| 41 | 2 | 2 | ,142 | 2 | ,919 | 3,901 | 1 | ,081 | 7,013 | 1,528 | -,212 |
| 42 | 2 | 2 | ,439 | 2 | ,999 | 1,649 | 1 | ,001 | 14,250 | 2,264 | -1,397 |
| 43 | 2 | 2 | ,104 | 2 | ,908 | 4,524 | 1 | ,092 | 7,342 | 4,455 | 1,036 |
| 44 | 2 | 2 | ,746 | 2 | ,997 | ,587 | 1 | ,003 | 10,417 | 3,948 | -,271 |
| 45 | 2 | 2 | ,599 | 2 | ,999 | 1,025 | 1 | ,001 | 13,708 | 4,383 | -,540 |
| 46 | 2 | 2 | ,687 | 2 | ,993 | ,751 | 1 | ,007 | 8,799 | 2,583 | -,552 |
| 47 | 2 | 2 | ,878 | 2 | 1,000 | ,260 | 1 | ,000 | 15,412 | 3,764 | -1,181 |
| 48 | 2 | 2 | ,748 | 2 | 1,000 | ,581 | 1 | ,000 | 17,504 | 3,890 | -1,407 |
| 49 | 2 | 2 | ,881 | 2 | 1,000 | ,253 | 1 | ,000 | 15,559 | 3,679 | -1,240 |
| 50 | 2 | 2 | ,746 | 2 | ,992 | ,586 | 1 | ,008 | 8,562 | 3,668 | -,095 |
| 51 | 2 | 2 | ,108 | 2 | ,989 | 4,452 | 1 | ,011 | 11,607 | 4,993 | ,578 |
| 52 | 2 | 2 | ,894 | 2 | ,994 | ,225 | 1 | ,006 | 8,848 | 3,207 | -,386 |
| 53 | 2 | 2 | ,510 | 2 | ,999 | 1,346 | 1 | ,001 | 14,511 | 4,539 | -,551 |
| 54 | 2 | 2 | ,344 | 2 | 1,000 | 2,134 | 1 | ,000 | 23,308 | 3,396 | -2,276 |
| 55 | 2 | 2 | ,656 | 2 | 1,000 | ,843 | 1 | ,000 | 17,298 | 4,249 | -1,186 |
| 56 | 2 | 2 | ,962 | 2 | ,999 | ,077 | 1 | ,001 | 13,313 | 3,666 | -,921 |
| 57 | 2 | 2 | ,546 | 2 | 1,000 | 1,210 | 1 | ,000 | 16,866 | 4,499 | -,958 |

Продолжение таблицы А. 2

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 58 | 2 | 2 | ,618 | 2 | 1,000 | ,963 | 1 | ,000 | 14,660 | 4,383 | -,699 |
| 59 | 2 | 2 | ,821 | 2 | ,991 | ,395 | 1 | ,009 | 7,967 | 3,119 | -,258 |
| 60 | 2 | 2 | ,813 | 2 | ,991 | ,415 | 1 | ,009 | 8,052 | 3,347 | -,174 |
| 61 | 2 | 2 | ,722 | 2 | ,983 | ,651 | 1 | ,017 | 7,035 | 3,215 | -,032 |
| 62 | 2 | 2 | ,583 | 2 | 1,000 | 1,079 | 1 | ,000 | 16,441 | 4,441 | -,936 |
| 63 | 2 | 2 | ,464 | 2 | 1,000 | 1,535 | 1 | ,000 | 16,761 | 4,648 | -,830 |
| 64 | 2 | 2 | ,096 | 2 | ,962 | 4,678 | 1 | ,038 | 9,407 | 4,769 | ,866 |
| 65 | 2 | 2 | ,330 | 2 | ,999 | 2,219 | 1 | ,001 | 14,957 | 2,082 | -1,493 |
| 66 | 2 | 2 | ,815 | 2 | ,998 | ,409 | 1 | ,002 | 11,478 | 2,782 | -,944 |
| 67 | 2 | 2 | ,561 | 2 | 1,000 | 1,156 | 1 | ,000 | 20,436 | 3,895 | -1,774 |
| 68 | 2 | 2 | ,894 | 2 | ,994 | ,224 | 1 | ,006 | 8,853 | 3,211 | -,385 |
| 69 | 2 | 2 | ,521 | 2 | 1,000 | 1,303 | 1 | ,000 | 20,216 | 4,272 | -1,562 |
| 70 | 2 | 2 | ,073 | 2 | ,996 | 5,223 | 1 | ,004 | 14,662 | 1,185 | -1,344 |
| 71 | 2 | 2 | ,667 | 2 | ,986 | ,811 | 1 | ,014 | 7,563 | 3,593 | ,066 |
| 72 | 2 | 2 | ,438 | 2 | ,998 | 1,653 | 1 | ,002 | 12,372 | 2,165 | -1,143 |
| 73 | 2 | 2 | ,473 | 2 | ,997 | 1,499 | 1 | ,003 | 11,077 | 2,192 | -,953 |
| 74 | 2 | 2 | ,852 | 2 | ,993 | ,321 | 1 | ,007 | 8,402 | 3,081 | -,353 |
| 75 | 2 | 2 | ,337 | 2 | ,934 | 2,174 | 1 | ,066 | 5,702 | 3,724 | ,625 |
| 76 | 2 | 2 | ,890 | 2 | 1,000 | ,232 | 1 | ,000 | 15,135 | 3,762 | -1,142 |
| 77 | 2 | 2 | ,286 | 2 | 1,000 | 2,505 | 1 | ,000 | 25,115 | 4,274 | -2,141 |
| 78 | 2 | 2 | ,903 | 2 | ,999 | ,203 | 1 | ,001 | 12,558 | 3,022 | -1,047 |
| 79 | 2 | 2 | ,557 | 2 | 1,000 | 1,171 | 1 | ,000 | 16,926 | 2,766 | -1,686 |
| 80 | 2 | 2 | ,683 | 2 | 1,000 | ,763 | 1 | ,000 | 14,846 | 4,282 | -,800 |
| 81 | 2 | 2 | ,929 | 2 | ,999 | ,148 | 1 | ,001 | 13,310 | 3,790 | -,858 |
| 82 | 2 | 2 | ,356 | 2 | 1,000 | 2,065 | 1 | ,000 | 19,743 | 4,807 | -1,149 |
| 83 | 2 | 2 | ,832 | 2 | ,991 | ,369 | 1 | ,009 | 8,114 | 3,098 | -,293 |
| 84 | 2 | 2 | ,820 | 2 | 1,000 | ,397 | 1 | ,000 | 13,990 | 2,966 | -1,264 |
| 85 | 2 | 2 | ,181 | 2 | ,870 | 3,416 | 1 | ,130 | 5,468 | 1,780 | ,058 |
| 86 | 2 | 2 | ,495 | 2 | 1,000 | 1,408 | 1 | ,000 | 20,408 | 3,258 | -1,992 |
| 87 | 2 | 2 | ,332 | 2 | ,879 | 2,206 | 1 | ,121 | 4,411 | 2,428 | ,300 |
| 88 | 2 | 2 | ,441 | 2 | ,969 | 1,636 | 1 | ,031 | 6,784 | 2,274 | -,224 |
| 89 | 2 | 2 | ,511 | 2 | ,997 | 1,343 | 1 | ,003 | 10,906 | 4,306 | -,082 |
| 90 | 2 | 2 | ,422 | 2 | 1,000 | 1,723 | 1 | ,000 | 18,311 | 4,706 | -1,019 |
| 91 | 2 | 2 | ,122 | 2 | 1,000 | 4,207 | 1 | ,000 | 27,214 | 2,849 | -2,788 |

Продолжение таблицы А. 2

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 92 | 2 | 2 | ,020 | 2 | 1,000 | 7,818 | 1 | ,000 | 37,771 | 5,309 | -2,866 |
| 93 | 3 | 3 | ,767 | 2 | 1,000 | ,530 | 1 | ,000 | 53,056 | -4,825 | ,112 |
| 94 | 3 | 3 | ,705 | 2 | 1,000 | ,698 | 1 | ,000 | 52,702 | -4,855 | ,287 |
| 95 | 3 | 3 | ,574 | 2 | 1,000 | 1,111 | 1 | ,000 | 41,028 | -3,403 | -,906 |
| 96 | 3 | 3 | ,669 | 2 | 1,000 | ,803 | 1 | ,000 | 55,658 | -5,008 | ,102 |
| 97 | 3 | 3 | ,922 | 2 | 1,000 | ,163 | 1 | ,000 | 45,035 | -4,258 | ,224 |
| 98 | 3 | 3 | ,791 | 2 | 1,000 | ,469 | 1 | ,000 | 36,772 | -3,503 | ,052 |
| 99 | 3 | 3 | ,787 | 2 | 1,000 | ,480 | 1 | ,000 | 37,548 | -3,461 | -,202 |
| 100 | 3 | 3 | ,764 | 2 | 1,000 | ,539 | 1 | ,000 | 47,650 | -4,534 | ,462 |
| 101 | 3 | 3 | ,057 | 2 | 1,000 | 5,717 | 1 | ,000 | 24,635 | -1,817 | -,676 |
| 102 | 3 | 3 | ,804 | 2 | 1,000 | ,437 | 1 | ,000 | 53,787 | -4,809 | -,090 |
| 103 | 3 | 3 | ,789 | 2 | 1,000 | ,473 | 1 | ,000 | 40,018 | -3,544 | -,486 |
| 104 | 3 | 3 | ,608 | 2 | 1,000 | ,996 | 1 | ,000 | 58,511 | -5,147 | -,078 |
| 105 | 3 | 3 | ,689 | 2 | 1,000 | ,745 | 1 | ,000 | 35,243 | -3,291 | -,114 |
| 106 | 3 | 3 | ,839 | 2 | 1,000 | ,352 | 1 | ,000 | 46,454 | -4,415 | ,367 |
| 107 | 3 | 3 | ,619 | 2 | 1,000 | ,958 | 1 | ,000 | 59,014 | -5,130 | -,227 |
| 108 | 3 | 3 | ,490 | 2 | 1,000 | 1,428 | 1 | ,000 | 34,188 | -3,001 | -,485 |
| 109 | 3 | 3 | ,982 | 2 | 1,000 | ,036 | 1 | ,000 | 44,151 | -3,997 | -,271 |
| 110 | 3 | 3 | ,630 | 2 | 1,000 | ,923 | 1 | ,000 | 34,417 | -3,461 | ,500 |
| 111 | 3 | 3 | ,171 | 2 | 1,000 | 3,534 | 1 | ,000 | 41,242 | -4,289 | 1,709 |
| 112 | 3 | 3 | ,124 | 2 | 1,000 | 4,168 | 1 | ,000 | 22,313 | -2,189 | ,391 |
| 113 | 3 | 3 | ,450 | 2 | 1,000 | 1,599 | 1 | ,000 | 54,672 | -4,244 | -1,427 |
| 114 | 3 | 3 | ,836 | 2 | 1,000 | ,359 | 1 | ,000 | 44,137 | -4,252 | ,425 |
| 115 | 3 | 3 | ,608 | 2 | 1,000 | ,995 | 1 | ,000 | 55,519 | -5,049 | ,271 |
| 116 | 3 | 3 | ,783 | 2 | 1,000 | ,488 | 1 | ,000 | 53,678 | -4,832 | -,002 |
| 117 | 3 | 3 | ,825 | 2 | 1,000 | ,384 | 1 | ,000 | 40,472 | -3,956 | ,422 |
| 118 | 3 | 3 | ,169 | 2 | 1,000 | 3,557 | 1 | ,000 | 29,447 | -2,353 | -,729 |
| 119 | 3 | 3 | ,581 | 2 | 1,000 | 1,087 | 1 | ,000 | 59,965 | -5,193 | -,232 |
| 120 | 3 | 3 | ,862 | 2 | 1,000 | ,297 | 1 | ,000 | 50,921 | -4,650 | ,056 |
| 121 | 3 | 3 | ,512 | 2 | 1,000 | 1,340 | 1 | ,000 | 48,687 | -4,710 | ,849 |
| 122 | 3 | 3 | ,928 | 2 | 1,000 | ,149 | 1 | ,000 | 50,213 | -4,536 | -,116 |
| 123 | 3 | 3 | ,899 | 2 | 1,000 | ,212 | 1 | ,000 | 51,536 | -4,613 | -,173 |
| 124 | 3 | 3 | ,570 | 2 | 1,000 | 1,125 | 1 | ,000 | 55,419 | -5,070 | ,367 |
| 125 | 3 | 3 | ,725 | 2 | 1,000 | ,643 | 1 | ,000 | 36,532 | -3,628 | ,441 |
| 126 | 3 | 3 | ,844 | 2 | 1,000 | ,340 | 1 | ,000 | 53,320 | -4,593 | -,547 |

Продолжение таблицы А. 2

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 127 | 3 | 3 | ,962 | 2 | 1,000 | ,077 | 1 | ,000 | 42,260 | -3,970 | ,043 |
| 128 | 3 | 3 | ,741 | 2 | 1,000 | ,601 | 1 | ,000 | 55,155 | -4,920 | -,055 |
| 129 | 3 | 3 | ,712 | 2 | 1,000 | ,679 | 1 | ,000 | 48,741 | -4,628 | ,507 |
| 130 | 3 | 3 | ,699 | 2 | 1,000 | ,715 | 1 | ,000 | 56,787 | -4,998 | -,159 |
| 131 | 3 | 3 | ,996 | 2 | 1,000 | ,009 | 1 | ,000 | 44,918 | -4,079 | -,225 |
| 132 | 3 | 3 | ,365 | 2 | 1,000 | 2,017 | 1 | ,000 | 49,605 | -4,825 | 1,085 |
| 133 | 3 | 3 | ,455 | 2 | 1,000 | 1,573 | 1 | ,000 | 64,029 | -5,318 | -,630 |
| 134 | 3 | 3 | ,027 | 2 | 1,000 | 7,256 | 2 | ,000 | 26,068 | -1,680 | -1,234 |
| 135 | 3 | 3 | ,939 | 2 | 1,000 | ,126 | 1 | ,000 | 47,650 | -4,168 | -,521 |
| 136 | 3 | 3 | ,533 | 2 | 1,000 | 1,257 | 1 | ,000 | 32,664 | -3,335 | ,601 |
| 137 | 3 | 3 | ,828 | 2 | 1,000 | ,378 | 1 | ,000 | 45,907 | -3,921 | -,735 |
| 138 | 3 | 3 | ,735 | 2 | 1,000 | ,616 | 1 | ,000 | 56,711 | -4,899 | -,409 |
| 139 | 3 | 3 | ,708 | 2 | 1,000 | ,692 | 1 | ,000 | 35,085 | -3,349 | ,049 |
| 140 | 3 | 3 | ,366 | 2 | 1,000 | 2,008 | 1 | ,000 | 33,455 | -2,825 | -,661 |
| 141 | 3 | 3 | ,922 | 2 | 1,000 | ,162 | 1 | ,000 | 45,017 | -4,257 | ,224 |
| 142 | 3 | 3 | ,703 | 2 | 1,000 | ,705 | 1 | ,000 | 55,990 | -4,983 | -,042 |
| 143 | 3 | 3 | ,955 | 2 | 1,000 | ,093 | 1 | ,000 | 41,534 | -3,883 | -,024 |
| 144 | 3 | 3 | ,940 | 2 | 1,000 | ,124 | 1 | ,000 | 42,025 | -3,989 | ,146 |
| 145 | 3 | 3 | ,486 | 2 | 1,000 | 1,443 | 1 | ,000 | 56,755 | -4,463 | -1,326 |
| 146 | 3 | 3 | ,815 | 2 | 1,000 | ,409 | 1 | ,000 | 38,584 | -3,774 | ,350 |
| 147 | 3 | 3 | ,949 | 2 | 1,000 | ,105 | 1 | ,000 | 42,691 | -3,860 | -,306 |
| 148 | 3 | 3 | ,134 | 2 | 1,000 | 4,020 | 1 | ,000 | 69,352 | -4,998 | -1,984 |
| 149 | 3 | 3 | ,513 | 2 | 1,000 | 1,334 | 1 | ,000 | 59,310 | -5,269 | ,131 |
| 150 | 3 | 3 | ,365 | 2 | 1,000 | 2,014 | 1 | ,000 | 56,616 | -4,305 | -1,577 |
| 151 | 3 | 3 | ,505 | 2 | 1,000 | 1,365 | 1 | ,000 | 48,385 | -3,817 | -1,285 |
| 152 | 3 | 3 | ,822 | 2 | 1,000 | ,393 | 1 | ,000 | 41,988 | -4,091 | ,458 |
| 153 | 3 | 3 | ,227 | 2 | 1,000 | 2,967 | 1 | ,000 | 63,987 | -4,733 | -1,788 |
| 154 | 3 | 3 | ,568 | 2 | 1,000 | 1,132 | 1 | ,000 | 55,683 | -4,460 | -1,184 |
| 155 | 3 | 3 | ,278 | 2 | 1,000 | 2,560 | 1 | ,000 | 26,745 | -2,596 | ,206 |
| 156 | 3 | 3 | ,592 | 2 | 1,000 | 1,050 | 1 | ,000 | 35,587 | -3,164 | -,435 |
| 157 | 3 | 3 | ,731 | 2 | 1,000 | ,626 | 1 | ,000 | 56,716 | -4,833 | -,571 |
| 158 | 3 | 3 | ,574 | 2 | 1,000 | 1,109 | 1 | ,000 | 53,694 | -4,981 | ,484 |
| 159 | 3 | 3 | ,937 | 2 | 1,000 | ,130 | 1 | ,000 | 48,584 | -4,245 | -,515 |

***. Неправильно классифицированное наблюдение*

# ПРИЛОЖЕНИЕ Б

Таблица Б.1

Кодировочная таблица для определения систолической дисфункции миокарда у подростков

| № | Показатель | Описание | Код показателя |
|---|---|---|---|
| 1 | diagnoz | Диагноз | 1- миокардиофиброз<br>2- дизпластическая кардимиопатия<br>3- гипертрофическая кардиомипатия<br>4- дилятационная кардимиопатия<br>5- аритмия<br>6- синдром перенапряжения миокарда<br>7- коарктация аорты<br>8- стеноз аорты<br>9- стеноз легочной артерии<br>10- дефект межжелудочковой перегородки<br>11- дефект межпредсердной перегродки<br>12- сочетанные пороки<br>13- тетрада Фалло<br>14- здоров |
| 2 | sn | Сердечная недостаточность | 1- без сердечной недостаточности<br>2- 1 стадия сердечной недостаточности<br>3- 2 стадия сердечной недостаточности<br>4- контрольная группа |
| 3 | raznica | Физическая нагрузка | 1- адекватная реакция на физическую нагрузку<br>2- неадекватная реакция а физическую нагрузку<br>3- норма |
| 4 | pol | Пол | 1- м<br>2- ж |
| 5 | im_srb | ИММУНОЛОГИЯ С-реактивный белок | 0 – отрицательный<br>1 - положительный ( определяется) |
| 6 | im_asl_o | ИММУНОЛОГИЯ антистрептолизин-О; абс.ед. | 0 – норма ( 0 – 200)<br>1 – более 200 |

# ПРИЛОЖЕНИЕ В

Таблица В.1

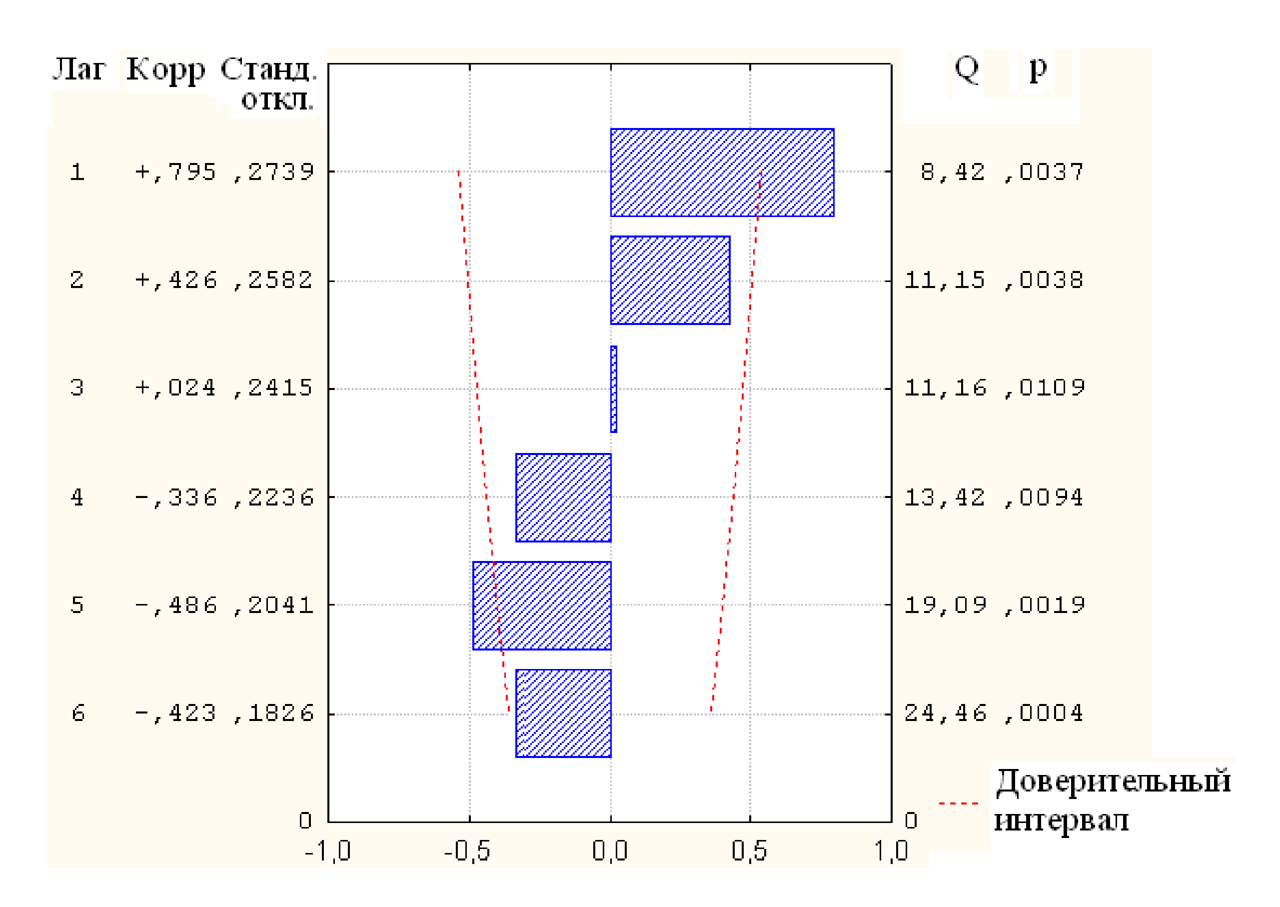
Таблица сопряженности для критерия Хосмера-Лемешова

| | | Группы = без ишемии миокарда | | Группы = ишемическая дилятационная кардиомиопатия | | |
|---|---|---|---|---|---|---|
| | | Наблюденные ($O_j$) | Ожидаемые ($E_j$) | Наблюденные ($O_j$) | Ожидаемые ($E_j$) | Всего ($n_j$) |
| Шаг 1 | 1 | 5 | 4,219 | 0 | ,781 | 5 |
| | 2 | 12 | 15,226 | 9 | 5,774 | 21 |
| | 3 | 9 | 5,162 | 3 | 6,838 | 12 |
| | 4 | 0 | 1,095 | 6 | 4,905 | 6 |
| | 5 | 0 | ,276 | 9 | 8,724 | 9 |
| | 6 | 0 | ,020 | 8 | 7,980 | 8 |
| | 7 | 0 | ,002 | 9 | 8,998 | 9 |
| | 8 | 0 | ,000 | 9 | 9,000 | 9 |
| | 9 | 0 | ,000 | 8 | 8,000 | 8 |
| Шаг 2 | 1 | 9 | 8,500 | 0 | ,500 | 9 |
| | 2 | 8 | 7,387 | 1 | 1,613 | 9 |
| | 3 | 7 | 6,271 | 2 | 2,729 | 9 |
| | 4 | 0 | 3,289 | 9 | 5,711 | 9 |
| | 5 | 2 | ,535 | 7 | 8,465 | 9 |
| | 6 | 0 | ,018 | 10 | 9,982 | 10 |
| | 7 | 0 | ,000 | 9 | 9,000 | 9 |
| | 8 | 0 | ,000 | 9 | 9,000 | 9 |
| | 9 | 0 | ,000 | 9 | 9,000 | 9 |
| | 10 | 0 | ,000 | 5 | 5,000 | 5 |
| Шаг 3 | 1 | 9 | 8,997 | 0 | ,003 | 9 |
| | 2 | 10 | 9,798 | 0 | ,202 | 10 |
| | 3 | 6 | 6,290 | 4 | 3,710 | 10 |
| | 4 | 1 | ,915 | 8 | 8,085 | 9 |
| | 5 | 0 | ,000 | 9 | 9,000 | 9 |
| | 6 | 0 | ,000 | 6 | 6,000 | 6 |
| | 7 | 0 | ,000 | 34 | 34,000 | 34 |
| Шаг 4 | 1 | 9 | 9,000 | 0 | ,000 | 9 |
| | 2 | 10 | 9,918 | 0 | ,082 | 10 |
| | 3 | 6 | 6,672 | 3 | 2,328 | 9 |
| | 4 | 1 | ,411 | 9 | 9,589 | 10 |
| | 5 | 0 | ,000 | 7 | 7,000 | 7 |
| | 6 | 0 | ,000 | 42 | 42,000 | 42 |

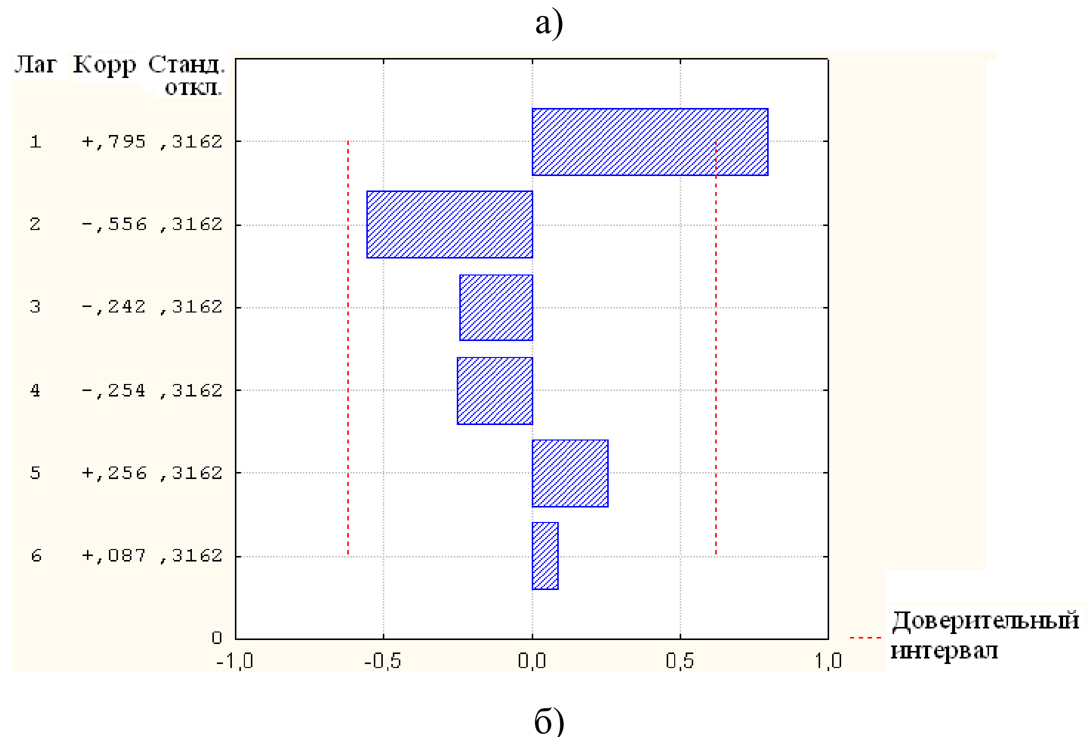
| Итерация | | -2 Log Правдоподобие | Коэффициенты | | | | |
|---|---|---|---|---|---|---|---|
| | | | Constant | САА | ППТ | ИЛ8 | ЛПВП |
| Шаг 1 | 1 | 99,840 | ,501 | ,320 | | | |
| | 2 | 80,335 | -,150 | 1,132 | | | |
| | 3 | 66,404 | -,795 | 2,330 | | | |
| | 4 | 58,004 | -1,432 | 3,923 | | | |
| | 5 | 54,502 | -1,961 | 5,577 | | | |
| | 6 | 53,657 | -2,308 | 6,809 | | | |
| | 7 | 53,594 | -2,426 | 7,255 | | | |
| | 8 | 53,593 | -2,437 | 7,295 | | | |
| | 9 | 53,593 | -2,437 | 7,295 | | | |
| Шаг 2 | 1 | 91,945 | -4,213 | ,278 | 2,514 | | |
| | 2 | 71,381 | -8,221 | ,996 | 4,409 | | |
| | 3 | 55,973 | -11,916 | 2,198 | 6,057 | | |
| | 4 | 46,433 | -14,951 | 3,868 | 7,321 | | |
| | 5 | 41,133 | -16,989 | 5,968 | 8,054 | | |
| | 6 | 38,668 | -19,586 | 8,376 | 9,088 | | |
| | 7 | 38,070 | -21,827 | 10,200 | 10,037 | | |
| | 8 | 38,029 | -22,631 | 10,828 | 10,385 | | |
| | 9 | 38,029 | -22,700 | 10,880 | 10,415 | | |
| | 10 | 38,029 | -22,700 | 10,880 | 10,415 | | |
| Шаг 3 | 1 | 89,849 | -4,471 | ,298 | 2,726 | -,002 | |
| | 2 | 67,861 | -8,904 | 1,040 | 4,878 | -,003 | |
| | 3 | 50,105 | -13,668 | 2,350 | 7,129 | -,004 | |
| | 4 | 38,692 | -18,655 | 4,217 | 9,482 | -,006 | |
| | 5 | 31,796 | -22,986 | 6,697 | 11,469 | -,008 | |
| | 6 | 26,226 | -28,760 | 10,777 | 14,025 | -,010 | |
| | 7 | 21,883 | -38,646 | 17,460 | 18,503 | -,013 | |
| | 8 | 19,066 | -54,641 | 27,470 | 25,888 | -,017 | |
| | 9 | 17,983 | -72,916 | 37,827 | 34,428 | -,022 | |
| | 10 | 17,804 | -84,136 | 44,008 | 39,676 | -,025 | |
| | 11 | 17,798 | -86,668 | 45,402 | 40,859 | -,026 | |
| | 12 | 17,798 | -86,764 | 45,456 | 40,904 | -,026 | |
| | 13 | 17,798 | -86,765 | 45,456 | 40,904 | -,026 | |
| Шаг 4 | 1 | 87,894 | -2,516 | ,315 | 2,609 | -,002 | -1,634 |
| | 2 | 66,268 | -6,905 | 1,022 | 4,980 | -,002 | -1,993 |
| | 3 | 48,269 | -11,794 | 2,332 | 7,516 | -,004 | -2,356 |
| | 4 | 36,946 | -16,668 | 4,181 | 10,165 | -,005 | -2,948 |
| | 5 | 30,402 | -21,219 | 6,554 | 12,399 | -,008 | -3,123 |
| | 6 | 24,474 | -27,586 | 10,778 | 14,972 | -,010 | -2,624 |
| | 7 | 18,896 | -38,958 | 18,668 | 20,393 | -,012 | -3,115 |
| | 8 | 14,276 | -60,809 | 32,726 | 31,253 | -,019 | -4,509 |
| | 9 | 11,840 | -90,575 | 50,581 | 46,292 | -,027 | -6,672 |
| | 10 | 11,048 | -118,135 | 67,132 | 60,278 | -,034 | -8,843 |
| | 11 | 10,943 | -132,178 | 75,793 | 67,405 | -,037 | -10,023 |
| | 12 | 10,940 | -134,738 | 77,442 | 68,699 | -,038 | -10,250 |
| | 13 | 10,940 | -134,818 | 77,496 | 68,739 | -,038 | -10,257 |
| | 14 | 10,940 | -134,818 | 77,496 | 68,739 | -,038 | -10,257 |

Рисунок В.1 – История итераций

# ПРИЛОЖЕНИЕ Г



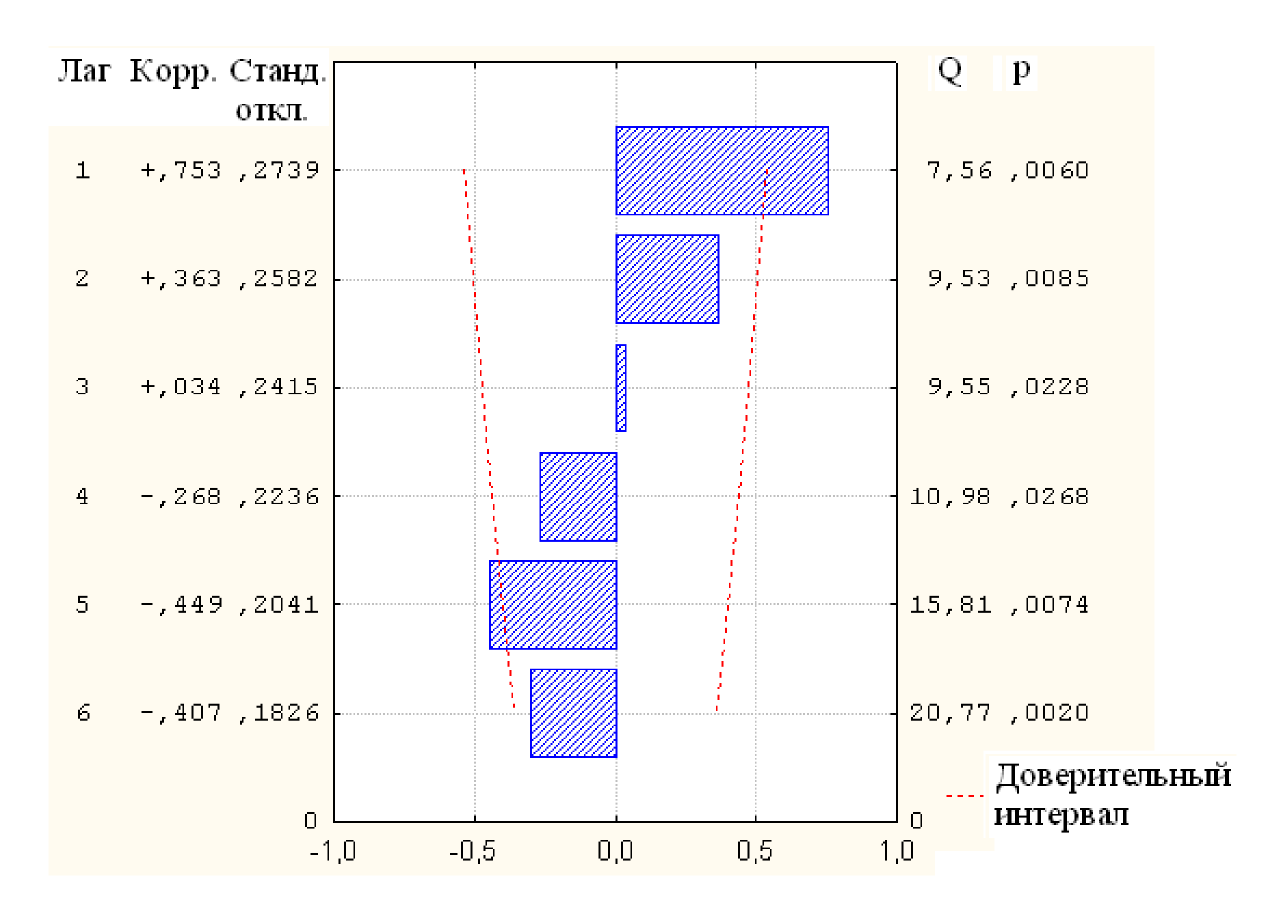

а)



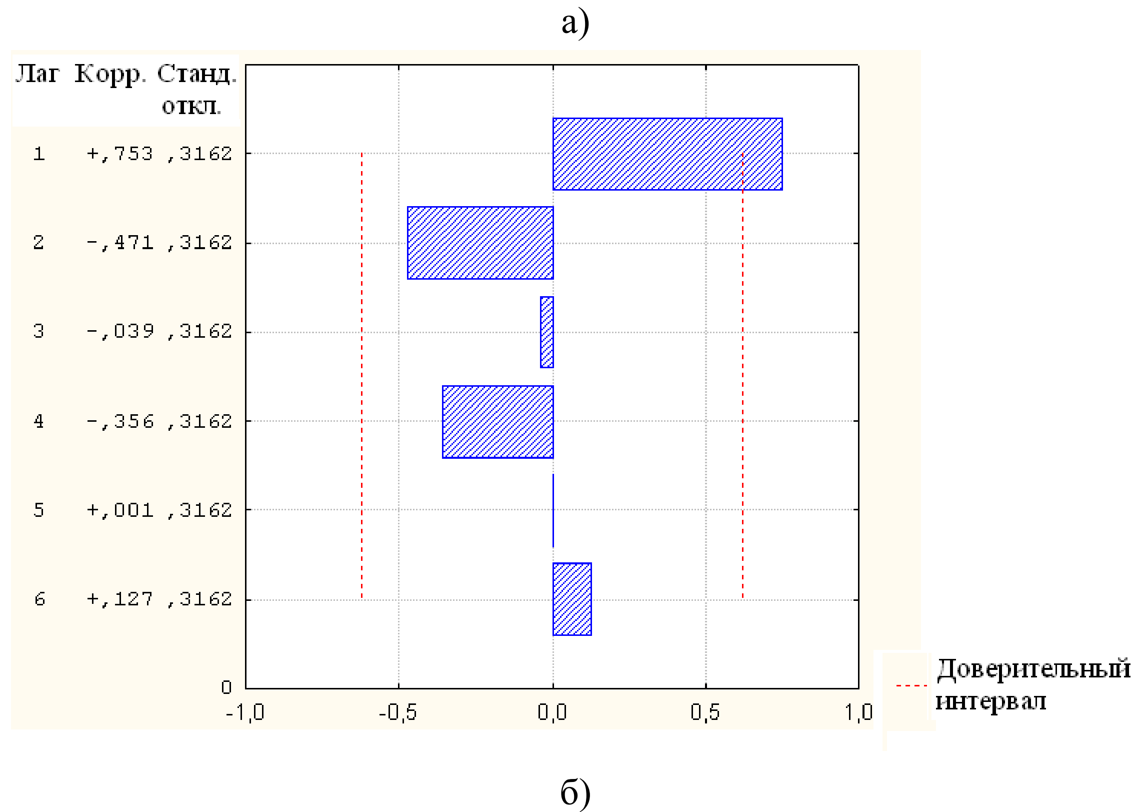

б)

Рисунок Г.1 –Автокорреляционные функции остатков для 2-го региона:

а - АКФ; б - ЧАКФ




а)




б)

Рисунок Г.2 – Автокорреляционные функции остатков для 3-го региона:

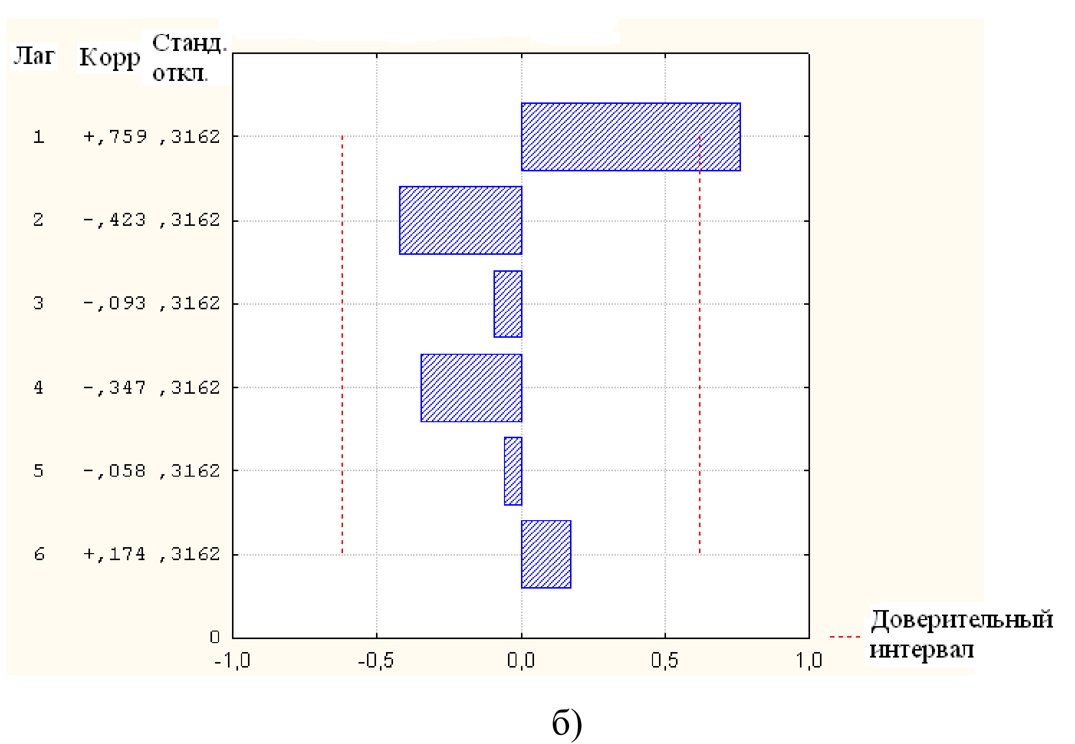
а - АКФ; б - ЧАКФ

| Лаг | Корр. | Станд. откл. | Q | p |
|---|---|---|---|---|
| 1 | +,759 | ,2739 | 7,69 | ,0056 |
| 2 | +,398 | ,2582 | 10,06 | ,0065 |
| 3 | +,076 | ,2415 | 10,16 | ,0173 |
| 4 | -,239 | ,2236 | 11,30 | ,0234 |
| 5 | -,449 | ,2041 | 16,14 | ,0065 |
| 6 | -,432 | ,1826 | 21,73 | ,0014 |

0 -1,0 -0,5 0,0 0,5 1,0 0

Доверительный интервал

а)




б)

Рисунок Г.3 – Автокорреляционные функции остатков для 4-го региона:

а - АКФ; б - ЧАКФ

Таблица Г.1

Параметры моделей ARMA для второго региона

| Модель Бокса-Дженкинса | Значения | | | Стандартная ошибка АКФ, SE | t-критерий | Значимость, P |
|---|---|---|---|---|---|---|
| | Константа | | 7713,684 | 176,831 | 43,622 | 0,0001 |
| ARMA (2, 1) | Коэффициенты AR | Лаг 1 | 1,677 | 0,122 | 13,702 | 0,0001 |
| | | Лаг 2 | -0,952 | 0,133 | -7,164 | 0,0001 |
| | Коэффициент MA | Лаг 1 | 0,993 | 37,845 | 0,026 | 0,0098 |
| ARMA (1, 2) | Константа | | 7732,531 | 756,785 | 10,218 | 0,0001 |
| | Коэффициент AR | Лаг 1 | 0,702 | 0,452 | 1,553 | 0,171 |
| | | Лаг 1 | -1,020 | 2,784 | -0,366 | 0,727 |
| | Коэффициенты MA | Лаг 2 | -0,056 | 0,693 | -0,081 | 0,938 |
| ARMA (2, 2) | Константа | | 7722,477 | 304,076 | 25,397 | 0,0001 |
| | Коэффициенты AR | Лаг 1 | 1,653 | 0,248 | 6,673 | 0,001 |
| | | Лаг 2 | -0,909 | 0,231 | -3,926 | 0,011 |
| | Коэффициенты MA | Лаг 1 | 0,721 | 29,883 | 0,024 | 0,982 |
| | | Лаг 2 | 0,263 | 8,599 | 0,031 | 0,977 |

Таблица Г.2

Параметры моделей ARMA для третьего региона

| Модель Бокса-Дженкинса | Значения | | | Стандартная ошибка АКФ, SE | t-критерий | Значимость, P |
|---|---|---|---|---|---|---|
| | Константа | | 6040,129 | 675,520 | 8,941 | 0,0001 |
| ARMA (2, 1) | Коэффициенты AR | Лаг 1 | 0,745 | 0,522 | 1,428 | 0,0203 |
| | | Лаг 2 | -0,116 | 0,549 | -0,212 | 0,0083 |
| | Коэффициент MA | Лаг 1 | -0,993 | 20,456 | -0,049 | 0,0430 |
| ARMA (1, 2) | Константа | | 6045,998 | 700,926 | 8,626 | 0,0001 |
| | Коэффициент AR | Лаг 1 | 0,650 | 0,554 | 1,173 | 0, 0285 |
| | | Лаг 1 | -0,891 | 1,718 | -0,519 | 0,623 |
| | Коэффициенты MA | Лаг 2 | 0,031 | 0,940 | 0,033 | 0,975 |
| ARMA (2, 2) | Константа | | 6001,893 | 556,726 | 10,781 | 0,0001 |
| | Коэффициенты AR | Лаг 1 | 1,558 | 0,937 | 1,662 | 0,0157 |
| | | Лаг 2 | -0,725 | 0,618 | -1,173 | 0,0294 |
| | Коэффициенты MA | Лаг 1 | 0,00032 | 57,444 | 0,0000005 | 0,0100 |
| | | Лаг 2 | 0,991 | 66,784 | 0,015 | 0,0890 |

Таблица Г.3

Параметры моделей ARMA для четвертого региона

| Модель Бокса-Дженкинса | Значения | | | Стандартная ошибка АКФ, SE | t-критерий | Значимость, P. |
|---|---|---|---|---|---|---|
| | Константа | | 3154,232 | 499,656 | 6,313 | 0,0010 |
| ARMA (2, 1) | Коэффициенты AR | Лаг 1 | 0,780 | 0,541 | 1,441 | 0,0200 |
| | | Лаг 2 | -0,084 | 0,575 | -0,146 | 0,0090 |
| | Коэффициент MA | Лаг 1 | -0,999 | 182,897 | -0,005 | 0,0060 |
| ARMA (1, 2) | Константа | | 3153,977 | 553,845 | 5,695 | 0,001 |
| | Коэффициент AR | Лаг 1 | 0,711 | 0,443 | 1,606 | 0,159 |
| | | Лаг 1 | -1,046 | 2,334 | -0,448 | 0,670 |
| | Коэффициенты MA | Лаг 2 | -0,099 | 0,659 | -0,150 | 0,886 |
| ARMA (2, 2) | Константа | | 3139,145 | 386,215 | 8,128 | 0,0001 |
| | Коэффициенты AR | Лаг 1 | 1,628 | 0,825 | 1,974 | 0,0105 |
| | | Лаг 2 | -0,779 | 0,594 | -1,312 | 0,0247 |
| | Коэффициенты MA | Лаг 1 | 0,005 | 52,787 | 0,0009 | 1,000 |
| | | Лаг 2 | 0,985 | 57,475 | 0,017 | 0,0070 |

## ЛИТЕРАТУРА


1. Доказательная медицина [Текст]. Т. 6. Клиническая фармакология. – 1999. – С. 3-5.

2. Скворцова, В.И. Современные методы доказательной медицины [Текст] / В.И. Скворцова // Журнал неврологии и психиатрии им. С.С. Корсакова. – 2002. – Т. 101, № 8. – С. 32-34.

3. Моисеев, С. Медицина, основанная на доказательствах [Текст] / С. Моисеев // Врач. – 2000. – № 12. – С. 4-8.

4. Заугольников, В.С. Идеи доказательной медицины - в учебный процесс и научные исследования [Текст] / В.С. Заугольников // Вятский медицинский вестник. – 2001. – № 1. – С. 44-45.

5. Кржечковская, В.Г. Заболевания сердечно-сосудистой системы детей и подростков [Текст] / В. Кржечковская, Р. Вахтангишвили. – М.: Феникс, 2006. – 508 с.

6. Воронков, Л.Г. Первичная профилактика сердечной недостаточности – один из приоритетов современной кардиологии [Текст] / Л.Г. Воронков // Укр. кардіологічний журнал. – 2004. – № 4. – С. 9-13.

7. Добрин, Б.Ю. Современные представления о механизмах начальных проявлений и прогрессирования сердечной недостаточности [Текст] / Б.Ю. Добрин, И.Е. Белая // Укр. кардіологічний журнал. – 2005. – № 6. – С. 143-150.

8. Алиев, Т.И. Основы моделирования дискретных систем [Текст] / Т.И. Алиев. – СПб.: СПбГУ ИТМО, 2009. – 363 с.

9. Дюк, В. А. Компьютерная психодиагностика [Текст] / В.А. Дюк. – С-Пб.: Питер, 1994. – 318 с.

10. Каримов, Р.Н. Обработка экспериментальной информации [Текст]. Ч. 3. Многомерный анализ : уч. пособие / Р.Н. Каримов. – Саратов: СГТУ, 2000. – 108 с.

11.Справочник по прикладной статистике [Текст]. В 2-х т. Т. 2 / Под ред. Э. Ллойда, У. Ледермана, С.А. Айвазяна, Ю.Н. Тюрина. – М.: Финансы и статистика, 1990. – 526 с.

12. Tests For Equality Of Variance (discriminant analysis algorithms) [Электронный ресурс]. – Режим доступа: \www/ URL: http://publib.boulder.ibm.com/infocenter/spssstat.

13. Лекции по SPSS [Электронный ресурс]. – Режим доступа: \www/ URL: http://docpsy.ru/lektsii/spss.html.

14. Наследов, А.Д. SPSS 19. Профессиональный статистический анализ данных. – СПб.: Питер, 2011. – 400 с.

15. Символ Кронекера [Электронный ресурс]. – Режим доступа: \www/ URL: http://ru.wikipedia.org/wiki/.

16. Разложение Холецкого [Электронный ресурс]. – Режим доступа: \www/ URL: http://ru.wikipedia.org/wiki/.

17. Основы дискриминантного анализа [Электронный ресурс]. – Режим доступа: \www/ URL: http://www.uran.donetsk.ua/~masters/2005/kita/ kapustina/ library/ discr_an.htm.

18. Бых, А.И. Использование дискриминантного анализа для диагностики хронической сердечной недостаточности у подростков [Текст] / А.И. Бых, Е.В. Высоцкая, А.П. Порван, Л.И. Рак, В.Г. Антоненко, Е.Е. Болибок, О.А. Сватенко // Вісник Національного технічного університету "Харківський політехнічний інститут". Збірник наукових праць. Тематичний випуск: Інформатика і моделювання. – Харків: НТУ "ХПІ", 2010. – №31. – С. 16-22.

19. Дискриминантный анализ: Методические указания к лабораторной работе [Текст] / Сост. В. А. Марченко; Р. Н. Каримов, Волгоград. гос. техн. ун-т. – Волгоград, 2004. – 25 с.

20. Ким, Дж.-О. Факторный, дискриминантный и кластерный анализ [Текст]: пер. с англ. / Дж.-О. Ким, Ч. У. Мьюллер, У. Р. Клекка и др.; Под ред. И. С. Енюкова. — М.: Финансы и статистика, 1989.— 215 с.

21. Выслоух, С.П. Применение методов дискриминантного анализа при технологическом проектировании [Электронный ресурс] / С.П. Выслоух. – Режим доступа: \www/ URL: http://users.kpi.kharkov.ua/cutting/articles/000028/000028.pdf/.

22. Шкала Чеддока [Электронный ресурс]. – Режим доступа: \www/ URL: http://math.semestr.ru/corel/cheddok.php.

23. Гребешкова, И.А. Методика построения и использования дискриминантной модели для налоговой диагностики [Электронный ресурс] / И.А. Гребешкова. – Режим доступа: \www/ URL: http://sevntu.com.ua/jspui/bitstream/123456789/494/1/vestnik92.2008.econom.40-43.pdf.

24. Буреева, Н.Н. Многомерный статистический анализ с использованием ППП “STATISTICA”. [Текст] : учебно-методический материал по программе повышения квалификации «Применение программных средств в научных исследованиях и преподавании математики и механики». – Нижний Новгород, 2007. – 112 с.

25. Айвазян, С. А. Прикладная статистика: классификация и снижение размерности: справ.изд. [Текст] / С. А. Айвазян, В. М. Бухштабер, И. С. Енюков, Л. Д. Мешалкин. – М.: Финансы и статистика, 1989. – С. 83-142.

26. Болч, Б. Многомерные статистические методы для экономики [Текст] / Б. Болч, К. Дж Чуань. – М.: Статистика, 1979. –С. 220-228.

27. Кади, Дж. Количественные методы в экономике [Текст] / Дж. Кади. – М.: Прогресс, 1977. – 246 с.

28. Высоцкая, Е.В. Методика определения систолической дисфункции миокарда у подростков [Текст] / Е.В. Высоцкая, О.А. Сватенко, Л.И. Рак, А.П. Порван // Восточно-Европейский журнал передовых технологий. – 2012. – Т. 1, №3(55). - С. 27-31.

29. Visotska, O.V. Diagnosis of systolic dysfunction in adolescents, [Текст] / O.V. Visotska, L.I. Rak, A.P. Porvan, O.A. Svatenko // Nauka I Studia. – Pizemisl, 2012. – № 3 (48). – pp. 64-72.

30. Дискриминантный анализ [Электронный ресурс]. – Режим доступа: \www/ URL: http://advschool.ru/articles/ article1002.htm.

31. Халафян, А.А. Дискриминантный анализ в определени кредитоспособности предприятий [Электронный ресурс] / А.А. Халафян, Е.Ю. Пелипенко. – Режим доступа: \www/ URL: http://www.statlab.kubsu.ru/sites/default/files/pdf/st4.pdf.

32. Дискриминантный анализ [Электронный ресурс]. – Режим доступа: \www/ URL: http://cdo.bseu.by/stat1/lab_4.htm.

33. Жученко, Ю. М. Информационные технологии в биологии и химии: лабораторный практикум для студентов вузов [Текст] / Ю. М. Жученко; М-во образования РБ, Гомельский гос. ун-т им. Ф. Скорины. – Гомель: ГГУим. Ф. Скорины, 2010.– 148 с.

34. Overall, J. Applied multivariate analysis[Текст] / J. Overall, C. Klett. – New York: MC Graw - Hill, 1972.

35. Лесовой, В.Н. Применение перитонеального диализа как метода коррекции уремического синдрома у пациентов с хронической болезнью

почек [Электронный ресурс] / В.Н. Лесовой, Н.М. Андоньева, Е. А. Гуц, А.В. Лесовая, М.Я. Дубовик. – Режим доступа: \www/ URL: http://archive.nbuv.gov.ua/portal/chem_biol/mmzh/2011_3/pdf/ 20110318.pdf.

36. Клиническое практическое руководство по адекватности гемодиализа. Инициативы качества диализных исходов НПФ США. [Текст]: пер. с англ. / А.Ю. Земченкова, ред. пер. Н.А. Томилина. – Тверь: ООО "Губернская медицина", 2000. – 190 с.

37. Elsner, D. How to diagnose and treat coronary artery disease in the uremic patient: an update. [Текст] / D. Elsner // Nephrol Dial Transplant. – 2001. – №16. – pp. 1103-1108.

38. Snowden, C.B. Amer. J. Epidem [Электронный ресурс] / C.B. Snowden, P.M. McNamara, R.J. Garrison eal. – 1982. – Vol. 115. – pp. 217-222. – Режим доступа: \www/ URL: http://www.medlinks.ru/bibl/pulm/21.htm.

39. Бинарная логистическая регрессия [Электронный ресурс]. – Режим доступа: \www/ URL: http://www.datuapstrade.lv/ rus/spss/section_16/4/.

40. Тимофеев, В.С. Модели бинарного выбора: оценка качества и интепретация коффициентов [Электронный ресурс] / В.С. Тимофеев, А.В. Большакова. – Режим доступа: \www/ URL: http://sbornik.infoterra.ru/dowload2.php?num=2-40&id=3.

41. Logistic regression [Электронный ресурс]. – Режим доступа: \www/ URL: http://www.bhsu.edu/instres/ logistic_regression.pdf.

42. Регрессионный анализ [Электронный ресурс]. – Режим доступа: \www/ URL: http://tempus2010.usm.md/SPSS10/16.pdf.

43. Research Methodology in Strategy and Management: Volume 5 [Электронный ресурс]. – Режим доступа: \www/ URL: http://books.google.com.ua.

44. Log-likelihood (multinomial logistic regression algorithms) [Электронный ресурс]. – Режим доступа: \www/ URL: http://pic.dhe.ibm.com/infocenter/spssstat/v21r0m0/index.jsp?topic=%2Fcom.ibm.spss.statistics.help%2Falg_nomreg_statistics_rsquare_nagelkerke.htm.

45. Sarkar, S. K. Validation and Performance Analysis of Binary Logistic Regression Model [Электронный ресурс] / Sohel Rana, Habshah Midi, and S. K. Sarkar. – Режим доступа: \www/ URL: http://www.wseas.us/e-library/conferences/2010/Penang/ EMEH/EMEH-09.pdf.

46. Goodness of Fit Measures [Электронный ресурс]. – Режим доступа: \www/ URL: http://www.strath.ac.uk/aer/materials/5furtherquantitativeresearchdesignandanalysis/unit6/goodnessoffitmeasures/.

47. Hosmer-Lemeshow goodness of fit test for Survey data [Электронный ресурс]. – Режим доступа: \www/ URL: http://www.amstat.org/sections/srms/Proceedings/y2003/Files/JSM2003-000744.pdf.

48. Lemeshow, S. The use of goodness-of-fit statistics in the development of logistic regression models [Текст] / S. Lemeshow and D. W. HosmerAmerican // Journal of Epidemiology. – Vol.115. – 1982. – pp. 92-106.

49. Мешкова, Т.А. Психогенетика [Электронный ресурс] / Т.А. Мешкова, Н.О. Николаева. – Режим доступа: \www/ URL: http://imp.rudn.ru/psychology/psychogenetic/index.html.

50. Дубницкая, Э.Б. Проблема резистентности при биполярных аффективно-бредовых расстройствах [Электронный ресурс] / Э.Б. Дубницкая // Психиатрия и психофармакотерапия. – Том 04, № 4. – 2002. – Режим доступа: \www/ URL: http://www.consilium-medicum.com/article/10570.

51. Спіріна, І.Д. Фактори ризику посилення частоти та збільшення тривалості госпіталізацій хворих органічними психічними розладами пізнього віку з супутньою соматичною патологією [Текст] / І.Д. Спіріна, С.Ф. Леонов, С.В. Рокутов, Т.Й. Шустерман, А.Є. Ніколенко, Т.В. Мозолюк // Журнал психиатрии и медицинской психологии. – № 2 (27). – 2011. – С. 18 – 22.

52. Суслов, В.И. Эконометрия [Текст] / В.И. Суслов, Н.М. Ибрагимов и др. – Новосибирск: Изд-во СО РАН, 2005. — 744 с.

53. Державна служба статистики України [Электронный ресурс]. – Режим доступа: \www/ URL: http://www.ukrstat.gov.ua/.

54. Гребенников, А.В. Моделирование сетевого трафика и прогнозирование с помощью модели ARIMA [Текст]/ А.В. Гребенников, Ю.А. Крюков, Д.В. Чернягин // Системный анализ в науке и образовании. Электронное научное издание. – 2011. – №1. – 11 с. – ISSN 2071-9612. – Режим доступа: \www/ URL: http:/www.sanse.ru/archive/19.

55. Горбацевич, В.В. Анализ и прогнозирование временных рядов [Текст] / В.В. Горбацевич. – М.: 2000. – Часть 2. – 26 с.

56. Бондарь, В.И. Современные методы анализа временных рядов [Текст] / В.И. Бондарь // Вторая Международная конференция "Системный анализ и информационные технологии". Том 1. – 2007. – С. 138-142.

57. Анализ временных рядов [Электронный ресурс]. – Режим доступа: \www/ URL: http://www.statsoft.ru/home/textbook/modules/sttimser.html.

58. Лемешко, Б.Ю. Проверка гипотез о математических ожиданиях и дисперсиях в задачах метрологии и контроля качества при вероятностных законах, отличающихся от нормального [Текст] / Б.Ю. Лемешко, С.С. Помадин // Метрология. – 2004. – № 3. – С. 3-15.

59. Лемешко, Б.Ю. Об устойчивости и мощности критериев проверки однородности средних [Текст] / Б.Ю. Лемешко, С.Б. Лемешко // Измерительная техника. – 2008. – № 9. – С.23-27.

60. Беркович, А.С. Исследование распределений статистик критериев тренда и случайности [Текст] / А.С. Беркович, Б.Ю. Лемешко, А.Е. Щеглов // Материалы X международной конференции “Актуальные проблемы электронного приборостроения” АПЭП-2010. – Новосибирск, 2010. – Т.6. – С. 13-17.

61. Лукашин, Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов [Текст] : учеб. пособие. / Ю.П. Лукашин. – М.: Финансы и статистика, 2003. – 416 с.

62. Варюхин, С.Е. Анализ модельных временных рядов с помощью надстройки «ARIMA» [Электронный ресурс] / С.Е. Варюхин. – Режим доступа: \www/ URL: http://www.hcxl.ru/examples/ar_TimeSeries/AR_examples_MTS.pdf.

63. Дуброва, Т. А. Статистические методы прогнозирования в экономике [Текст] : учебно-методический комплекс / Т. А. Дуброва, М. Ю. Архипова. – М.: МЭСИ, 2008. – 136 с.

Підп. до друку 02.08.2013р. Формат 60х84/16. Надруковано на різографі
Gestetner 6123CP. Ум.-друк. арк. 8,5. Наклад 300 прим. Зам. № 17-13.
Ціна договірна

---

Видавництво та друкарня "Технологічний Центр"
Свідоцтво про внесення суб'єкта видавничої справи ДК №4452 від 10.12.2012
Адреса: 61145, м. Харків, вул. Шатилова дача, 4

---